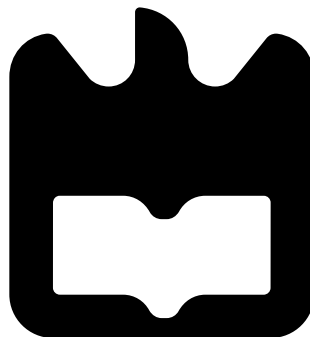




**Bruno Miguel Pinho  
Barbosa**

**Extração de características relacionadas com  
postura em situações de trabalho em grupo**







**Bruno Miguel Pinho  
Barbosa**

**Extração de características relacionadas com  
postura em situações de trabalho em grupo**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Informática, realizada sob a orientação científica de António José Ribeiro Neves, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro e coorientação de Sandra Cristina de Oliveira Soares, Professora Auxiliar do Departamento de Educação e Psicologia





**o júri / the jury**

presidente / president

**Professor Doutor Paulo Miguel de Jesus Dias**

Professor Auxiliar da Universidade de Aveiro

vogais / examiners committee

**Professor Doutor António José Ribeiro Neves**

Professor Auxiliar da Universidade de Aveiro (orientador)

**Professor Doutor Luís Filipe Pinto de Almeida Teixeira**

Professor Auxiliar do Departamento de Engenharia Informática da Faculdade de Engenharia da Universidade do Porto



**agradecimentos /**  
**acknowledgements**

Ao Professor António Neves, pelo excelente acompanhamento enquanto professor e orientador, mostrando sempre prontidão e vontade de ajudar. Foi sem dúvida uma das pessoas que mais me motivou para a escolha deste tema, criando oportunidades de aprofundamento de conhecimento únicas que jamais esquecerei. De realçar a relação Professor-Aluno que consegue criar, cativando ainda mais o meu interesse e de todos os seus alunos.

À Professora Sandra Soares, pela sua constante ajuda e dedicação do início ao fim deste projeto, pois sem ela não seria possível o seu desenvolvimento. Também de realçar a relação Professor-Aluno que consegue criar, deixando-me totalmente à vontade para esclarecer qualquer tipo de dúvidas ou trocar opiniões, numa área (Psicologia) na qual eu estaria um pouco inseguro inicialmente.

Aos meus pais e irmã, por todo o apoio dado durante estes 5 anos. Foram fundamentais para que conseguisse chegar até aqui. Sem eles nada disto teria sido possível e todas as experiências académicas e não académicas que tive não se teriam concretizado.

À Mara por estar sempre presente. Pelos momentos menos bons que tive de aturar, conseguindo fazer com que eu desse sempre a volta por cima. Sem este apoio teria sido muito mais difícil e não teria conseguido chegar aqui da forma que cheguei.

A todos os meus amigos e familiares que foram fundamentais durante todo o meu percurso. Desde os primeiros dias que o apoio de todos foi incrível. A forma como me receberam em Aveiro foi excelente, tornando a mudança radical de vida muito mais simples.



## Palavras-Chave

Estimativa de Postura; Imagem Digital; Emoções; Detecção do Esqueleto.

## Resumo

A expressão facial do ser humano é um veículo de excelência na comunicação de emoções. No entanto, o avanço no estudo das emoções mostra que as expressões ou posturas corporais têm um desempenho, no reconhecimento das emoções, semelhante ao das expressões faciais. Neste sentido, é necessário perceber, tecnologicamente, de que forma se podem extrair características da postura que permitam efetuar uma análise detalhada da mesma, percebendo as emoções associadas.

O pouco avanço tecnológico nesta área deve-se à dificuldade em associar uma emoção a determinada postura, uma vez que esta avaliação não é de um domínio binário, existindo uma enorme variedade de emoções entre, por exemplo, a tristeza, alegria, entre outras. Assim, surge a necessidade de criar um sistema que permita extrair características da postura do corpo humano e gerar dados que permitam efetuar uma avaliação dessa mesma postura.

Nesta dissertação, começa-se por efetuar uma contextualização com o problema existente, seguindo-se um levantamento do estado de arte relativo a sistemas computacionais existentes para deteção e classificação da postura humana, sensores de imagem e algoritmos para deteção do esqueleto. Depois disto, é apresentada a solução desenvolvida para a lacuna existente no tipo de sistemas apresentados, explicando-se todos os procedimentos efetuados, passo a passo. Por fim, são apresentados os resultados conseguidos no que diz respeito às características obtidas do corpo humano (deslocações, áreas e sobreposições), através de gráficos e *Heatmaps*, e tiradas as conclusões necessárias para o contexto apresentado.



**Keywords**

Pose Estimation; Digital Image; Emotions; Skeleton Detection.

**Abstract**

The facial expression of the human being is a vehicle of excellence in communicating emotions. However, advances in the study of emotions shows that the body expressions or body postures have a performance in the recognition of emotions similar to facial expressions. In this way, it is necessary to technologically understand how posture features can be extracted in order to perform a detailed analysis, perceiving associated emotions.

The lack of technological advance in this area is due to the fact that it is difficult to associate an emotion with a certain posture, since this evaluation is not of a binary domain and there is a huge variety of emotions, between, for example, sadness, joy, among others. Thus, there is a need to create a system that allows to extract characteristics of the posture of the human body and generate data that allow to make an evaluation of the given posture.

In this dissertation, a contextualizing of the existing problem is addressed, followed by a state-of-the-art survey of existing computer systems for human posture detection and classification, image sensors and algorithms for skeletal detection. After this, a solution for the existing gap in the type of presented systems is proposed, explaining all the procedures performed, step by step. Finally, the obtained results are presented regarding to the characteristics obtained from the human body (displacements, areas and overlaps), through graphs and heatmaps, and taking the necessary conclusions for the presented context.





# Conteúdo

<b>Conteúdo</b>	<b>i</b>
<b>Lista de Figuras</b>	<b>iii</b>
<b>Lista de Tabelas</b>	<b>vii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Estado Arte</b>	<b>3</b>
2.1 Sistemas computacionais existentes para deteção e classificação da postura humana . . . . .	3
2.2 Sensores de Imagem . . . . .	5
2.2.1 Sensores de imagem no espectro do visível . . . . .	6
2.2.2 Sensores especiais . . . . .	7
Térmicos . . . . .	8
Multi/Hiper Espectrais . . . . .	8
Distância . . . . .	9
2.3 Algoritmos para deteção de postura . . . . .	11
2.4 Instalação <i>OpenPose</i> . . . . .	14
<b>3 Cenário de Testes</b>	<b>17</b>
3.1 Descrição do Protocolo . . . . .	18
3.2 <i>Software</i> de aquisição de dados . . . . .	19
3.3 Sincronismo pós aquisição . . . . .	21
3.4 Divisão de fases experimentais . . . . .	24

<b>4</b>	<b><i>Software</i> de Detecção</b>	<b>25</b>
4.1	Região de Interesse . . . . .	25
4.2	Pontos de Interesse . . . . .	27
4.3	Área Ocupada . . . . .	27
4.4	Sobreposições . . . . .	30
4.5	Deslocamento . . . . .	32
<b>5</b>	<b>Resultados</b>	<b>37</b>
5.1	Deslocamento . . . . .	37
5.2	Área Ocupada . . . . .	38
5.3	Interação . . . . .	40
<b>6</b>	<b>Conclusão</b>	<b>45</b>
	<b>Bibliografia</b>	<b>49</b>
<b>A</b>	<b>Teclas de interação do <i>Software</i> desenvolvido.</b>	<b>55</b>
A.1	<i>Software</i> de sincronismo . . . . .	55
A.2	<i>Software</i> de divisão das fases experimentais . . . . .	55
<b>B</b>	<b>Gráficos e <i>Heatmaps</i> obtidos relativamente às diferentes experiências realizadas.</b>	<b>56</b>
B.1	Média de deslocamentos em experiências sem indução de conflito . . . . .	56
B.2	Média de deslocamentos em experiências com indução de conflito de tarefa . . .	57
B.3	Área ocupada em experiências sem indução de conflito . . . . .	58
B.4	Área ocupada em experiências com indução de conflito de tarefa . . . . .	59
B.5	<i>Heatmaps</i> reativos a experiências sem indução de conflito . . . . .	60
B.6	<i>Heatmaps</i> reativos a experiências com indução de conflito de tarefa . . . . .	61

# Lista de Figuras

1.1	Postura extraída do corpo humano em duas situações distintas. . . . .	2
2.1	Reconhecimento das ações humanas baseadas em primitivas da postura. . . . .	4
2.2	Matriz de pixels numa imagem com Largura $X$ e Altura $Y$ . . . . .	5
2.3	Gradiente associado a uma região de uma imagem RGB. . . . .	5
2.4	Imagem <i>grayscale</i> e matriz de intensidade respetiva. . . . .	6
2.5	Aplicação de um filtro de <i>Bayer</i> . . . . .	7
2.6	Exemplo de sensor CCD e CMOS. . . . .	8
2.7	Imagem térmica com respetiva escala de cores. . . . .	9
2.8	Imagem multiespectral e hiperespectral. . . . .	10
2.9	Câmara do tipo TOF ( <i>SwissRange SR4000</i> ) e imagem obtida pela mesma. . . . .	11
2.10	Câmara do tipo <i>Structured Light</i> ( <i>Kinect</i> ) e imagem obtida pela mesma. . . . .	11
2.11	Câmara do tipo <i>Stereo</i> ( <i>Zed</i> ) e imagem obtida pela mesma. . . . .	11
2.12	Configuração de <i>Hardware</i> de uma <i>Kinect</i> . . . . .	12
2.13	Deteção do esqueleto humano ( <i>Kinect Skeletal Tracker</i> ). . . . .	13
2.14	Deteção do esqueleto humano ( <i>OpenPose</i> ). . . . .	14
3.1	Organização do laboratório e esquema de posições. . . . .	18
3.2	Instanciação do objeto <i>VideoWriter</i> . . . . .	20
3.3	Diagrama de fluxo do <i>Software</i> de aquisição desenvolvido. . . . .	21
3.4	Média de armazenamento necessário por vídeo em cada abordagem. . . . .	22
3.5	Método de exclusão de frames. . . . .	23
4.1	Configuração da região de interesse de uma câmara. . . . .	25
4.2	Corte da região de interesse. . . . .	26

4.3	Diagrama de fluxo do <i>Software</i> de corte de região de interesse desenvolvido. . .	26
4.4	Ordem dos pontos de interesse definidos pela biblioteca <i>OpenPose</i> . . . . .	28
4.5	Excerto de um ficheiro de pontos de interesse gerado pela biblioteca <i>OpenPose</i> . . .	28
4.6	Ficheiro de configuração da câmara 2. . . . .	29
4.7	Divisão e atribuição de id's estáticos. . . . .	30
4.8	Representação gráfica de uma sobreposição. . . . .	32
4.9	Diagrama de <i>workflow</i> da deteção de sobreposições. . . . .	33
4.10	<i>ColorMap</i> utilizado. . . . .	33
4.11	Representação dos vetores associados às 13 partes do corpo em estudo. . . . .	34
4.12	Estrutura do ficheiro de deslocamentos e áreas. . . . .	34
4.13	Processo de obtenção do gráfico dos deslocamentos. . . . .	35
5.1	Divisão dos gráficos obtidos pelas duas fases da experiência. . . . .	38
5.2	Média de deslocamentos da pessoa com id 1, sem indução de conflito. . . . .	39
5.3	Média de deslocamentos da pessoa com id 2, sem indução de conflito. . . . .	39
5.4	Média de deslocamentos da pessoa com id 3, sem indução de conflito. . . . .	39
5.5	Média de deslocamentos da pessoa com id 4, sem indução de conflito. . . . .	39
5.6	Média de deslocamentos da pessoa com id 1, com indução de conflito. . . . .	40
5.7	Média de deslocamentos da pessoa com id 2, com indução de conflito. . . . .	40
5.8	Média de deslocamentos da pessoa com id 3, com indução de conflito. . . . .	40
5.9	Média de deslocamentos da pessoa com id 4, com indução de conflito. . . . .	40
5.10	Área ocupada pela pessoa com id 1, sem indução de conflito. . . . .	41
5.11	Área ocupada pela pessoa com id 2, sem indução de conflito. . . . .	41
5.12	Área ocupada pela pessoa com id 3, sem indução de conflito. . . . .	41
5.13	Área ocupada pela pessoa com id 4, sem indução de conflito. . . . .	41
5.14	Área ocupada pela pessoa com id 1, com indução de conflito. . . . .	42
5.15	Área ocupada pela pessoa com id 2, com indução de conflito. . . . .	42
5.16	Área ocupada pela pessoa com id 3, com indução de conflito. . . . .	42
5.17	Área ocupada pela pessoa com id 4, com indução de conflito. . . . .	42
5.18	<i>Heatmap</i> da câmara 1 numa experiência sem indução de conflito. . . . .	43
5.19	<i>Heatmap</i> da câmara 2 numa experiência sem indução de conflito. . . . .	43
5.20	<i>Heatmap</i> da câmara 3 numa experiência sem indução de conflito. . . . .	43

5.21	<i>Heatmap</i> da câmara 1 numa experiência com indução de conflito. . . . .	43
5.22	<i>Heatmap</i> da câmara 2 numa experiência com indução de conflito. . . . .	44
5.23	<i>Heatmap</i> da câmara 3 numa experiência com indução de conflito. . . . .	44
B.1	Média de deslocamentos da pessoa com id 1, sem indução de conflito. . . . .	56
B.2	Média de deslocamentos da pessoa com id 2, sem indução de conflito. . . . .	56
B.3	Média de deslocamentos da pessoa com id 3, sem indução de conflito. . . . .	56
B.4	Média de deslocamentos da pessoa com id 4, sem indução de conflito. . . . .	56
B.5	Média de deslocamentos da pessoa com id 1, com indução de conflito de tarefa. . . . .	57
B.6	Média de deslocamentos da pessoa com id 2, com indução de conflito de tarefa. . . . .	57
B.7	Média de deslocamentos da pessoa com id 3, com indução de conflito de tarefa. . . . .	57
B.8	Média de deslocamentos da pessoa com id 4, com indução de conflito de tarefa. . . . .	57
B.9	Área ocupada pela pessoa com id 1, sem indução de conflito. . . . .	58
B.10	Área ocupada pela pessoa com id 2, sem indução de conflito. . . . .	58
B.11	Área ocupada pela pessoa com id 3, sem indução de conflito. . . . .	58
B.12	Área ocupada pela pessoa com id 4, sem indução de conflito. . . . .	58
B.13	Área ocupada pela pessoa com id 1, com indução de conflito de tarefa. . . . .	59
B.14	Área ocupada pela pessoa com id 2, com indução de conflito de tarefa. . . . .	59
B.15	Área ocupada pela pessoa com id 3, com indução de conflito de tarefa. . . . .	59
B.16	Área ocupada pela pessoa com id 4, com indução de conflito de tarefa. . . . .	59
B.17	<i>Heatmap</i> da câmara 1 numa experiência sem indução de conflito. . . . .	60
B.18	<i>Heatmap</i> da câmara 2 numa experiência sem indução de conflito. . . . .	60
B.19	<i>Heatmap</i> da câmara 3 numa experiência sem indução de conflito. . . . .	60
B.20	<i>Heatmap</i> da câmara 1 numa experiência com indução de conflito. . . . .	61
B.21	<i>Heatmap</i> da câmara 2 numa experiência com indução de conflito. . . . .	61
B.22	<i>Heatmap</i> da câmara 3 numa experiência com indução de conflito. . . . .	61



# Lista de Tabelas

A.1	Teclas de interação do <i>Software</i> de sincronismo. . . . .	55
A.2	Teclas de interação do <i>Software</i> de divisão das fases experimentais. . . . .	55





# Capítulo 1

## Introdução

Há muito tempo que as expressões faciais estão fortemente associadas à percepção das emoções. No entanto, cada vez mais o corpo é um meio importante para a expressão emocional, tendo este um forte papel na comunicação emocional [1]. Investigações desenvolvidas acerca do comportamento e estudos de neuroimagem mostram que o reconhecimento de emoções em expressões corporais tem um desempenho semelhante ao das expressões faciais [2] [3].

Com o avanço dos estudos das emoções associadas à postura corporal é necessário perceber, tecnologicamente, como se podem extrair, de forma não invasiva, essas mesmas emoções. São imensas as áreas de aplicação deste tipo de problemas. Desde a educação, por exemplo em contexto de sala de aula, uma vez que este é um local associado a momentos de ansiedade e stress, devido a diversos fatores como desinteresse, desmotivação, elevado número de alunos e o seu mau comportamento [4], passando pela saúde mental e até mesmo situações de conflito entre um ou mais grupos de pessoas.

Pretende-se, assim, através da utilização de câmaras digitais e algoritmos que permitem extrair postura do ser humano, classificar determinadas posturas de forma a que sejam reconhecidas as emoções associadas a cada instante. A Figura 1.1 apresenta a postura extraída do corpo humano em duas situações distintas, na esquerda uma situação em que a pessoa se encontra pouco confiante e na direita uma situação de demonstração de confiança [5].

O objetivo principal deste projeto de mestrado consiste na criação de um sistema que permita reconhecer e classificar estas posturas, em vários grupos de pessoas, de forma dinâmica e eficaz, ou seja, sem que existam posturas pré-definidas como referido nos estudos já existentes [6] [7] [8]. Deste modo surgem assim as perguntas: “Como utilizar um PC e uma câmara para

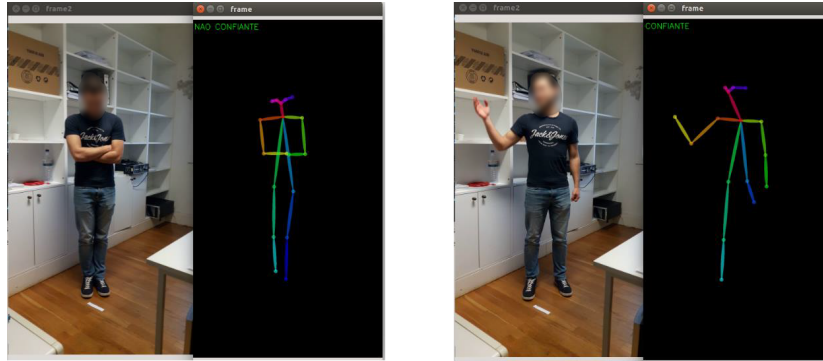


Figura 1.1: Postura extraída do corpo humano em duas situações distintas, na esquerda uma situação em que a pessoa se encontra pouco confiante e na direita uma situação de demonstração de confiança [5].

---

medir a postura corporal do corpo humano?” e “Como classificar cada postura como estando associada a determinada emoção?”.

Este projeto pretende contribuir para o avanço do estudo da postura corporal e das emoções associadas, uma vez que esta é uma área muito pouco explorada devido à dificuldade de avaliação das posturas consoante as suas emoções. Desta forma, a criação deste tipo de sistema torna-se um avanço notável nesta área científica. Neste sentido foram elaborados e publicados dois artigos científicos, [9] e [5], que dizem respeito a um levantamento de estado de arte contextualizado com este projeto e a um sistema de deteção e classificação de posturas num contexto individual, respetivamente. A cooperação com o Departamento de Educação e Psicologia, que será referida posteriormente, foi também fundamental para o sucesso deste sistema, nomeadamente na avaliação das emoções.

Durante esta dissertação será apresentado, no Capítulo 2, um levantamento do estado de arte acerca de sistemas computacionais existentes para deteção e classificação da postura humana, sensores de imagem e algoritmos para deteção de postura. De seguida, no Capítulo 3 será apresentado o cenário de testes, descrevendo o protocolo experimental, explicado a forma de aquisição de dados e o *Software* de aquisição desenvolvido, o sincronismo pós-aquisição e divisão dos vídeos de acordo com as fases da experiência. O Capítulo 4 apresentará o *Software* de deteção e extração de características desenvolvido, seguido dos resultados, apresentados no Capítulo 5. Por fim, no Capítulo 6, é feita uma conclusão acerca dos resultados obtidos e do sistema criado, incluindo sugestões de trabalho futuro.

## Capítulo 2

# Estado Arte

Durante este capítulo é apresentado um levantamento do estado da arte acerca de sistemas de visão computacional para detecção e classificação da postura humana que visam resolver o problema apresentado no capítulo anterior, incluindo o tipo de sensores existentes para a obtenção de imagens que alimentam este tipo de sistemas e o seu funcionamento, e algoritmos de detecção do esqueleto humano, dando ênfase aos mais utilizados, explicando os seus princípios. Por fim, será apresentado o processo de configuração e instalação de um computador com o *Software* de detecção de posturas, *OpenPose* [10].

### 2.1 Sistemas computacionais existentes para detecção e classificação da postura humana

Tem sido realizado algum trabalho no desenvolvimento de sistemas de detecção e avaliação das posturas do ser humano. No entanto, não existe nenhum que permita, de forma precisa, detetar e classificar essas mesmas posturas como estando associadas a emoções específicas. Esta escassez de sistemas deve-se à dificuldade de classificar uma postura, sendo que, na maioria dos casos, as pessoas adotam posturas nem muito contraídas, nem muito expansivas, tornando difícil perceber qual o seu estado emocional. Em [11], num contexto de sala de aula, os autores pretendem fornecer informações importantes ao professor, acerca da atenção da sua audiência. Este estudo foca-se principalmente na captura de dados através de um sistema de câmaras, sendo processados esses dados de forma a que sejam detetados os movimentos, a cabeça e a sua orientação e assim é possível obter os padrões de comportamento mais

significativos e fazer uso dos mesmos. No entanto, os resultados não mostraram uma relação direta entre os movimentos dos alunos e a sua atenção.

Em [12] é descrito um sistema de reconhecimento de ações humanas baseadas em primitivas da postura. Este sistema, à semelhança de [13], apenas se foca em perceber/classificar se uma pessoa corre, caminha, dança, etc. e não nas suas emoções. Numa fase de aprendizagem, estimam-se os parâmetros representativos da postura através de vídeos. Depois disso, já numa fase de classificação, o método é utilizado tanto para vídeos como para imagens estáticas. Neste sistema são identificados 3 problemas disjuntos no reconhecimento da ação humana: detecção de uma pessoa na imagem, reconhecimento da postura expressa e atribuição de uma categoria de ação à respetiva postura, sendo foco principal os últimos 2 pontos. Os resultados deste sistema são positivos, resultando numa alta precisão no reconhecimento de ações, de onde se retira que a postura do ser humano contém informações suficientes sobre a sua atividade. É referido ainda que, a junção de outras informações para além da postura, em conjunto com a mesma, permitem obter uma maior precisão no reconhecimento das atividades. A Figura 2.1 apresenta o processo de reconhecimento das ações humanas apresentado em [12].

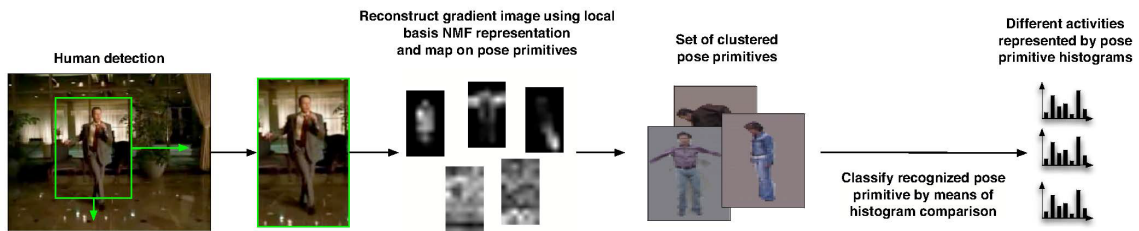


Figura 2.1: Processo de reconhecimento das ações humanas baseadas em primitivas da postura [12].

Em suma, nesta Secção verifica-se a existência de alguns sistemas de reconhecimento de postura com aplicações específicas. Contudo, não existe nenhum que permita reconhecer e classificar posturas de acordo com as emoções que transmitem.

## 2.2 Sensores de Imagem

Tipicamente, uma imagem digital é representada por uma matriz retangular de vetores escalares, sendo esta composta por um número finito de elementos numa determinada posição e com um determinado valor. Esses elementos são chamados de pixels [14]. Um pixel é a unidade mais pequena de uma imagem e tem associado a si um valor de intensidade e uma localização. Através da junção de muitos pixels e devido ao efeito de filtragem do olho humano, é possível criar ilusões como gradientes e sombreamento. A Figura 2.2 representa uma matriz de pixels relativos a uma imagem digital e a Figura 2.3 representa o gradiente de uma imagem *Red, Green and Blue* (RGB) através da junção de pixels.

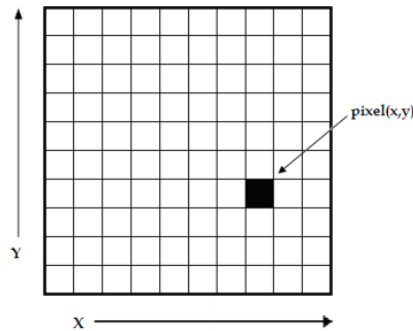


Figura 2.2: Representação de uma matriz de pixels numa imagem com Largura  $X$  e Altura  $Y$  [15].

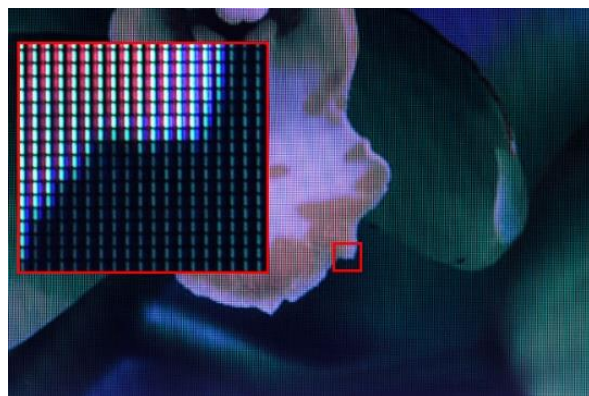


Figura 2.3: Gradiente associado a uma região de uma imagem RGB [16].

Os tipos de imagens mais comuns quando se fala em imagem digital são imagens *grayscale*

e RGB. Em imagens *grayscale*, o valor associado a cada pixel é preto, branco ou um tom de cinza, podendo este variar, para 8 bits por pixel, entre 0 e 255, em que 0 corresponde a preto e 255 a branco. Em imagens a cores cada pixel tem associado a si um valor de vermelho, verde e azul, que combinados em diferentes quantidades podem gerar qualquer cor. Os valores de vermelho, verde e azul variam também, para 8 bits por pixel, entre 0 e 255, sendo 0 a cor preta e 255 o máximo da cor respectiva. A Figura 2.4 representa uma imagem *grayscale* e a matriz de intensidade respectiva a determinada área da mesma [17].

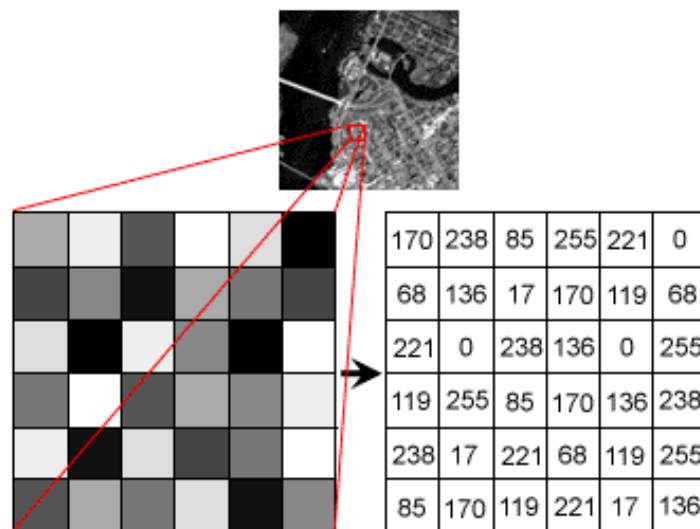


Figura 2.4: Imagem *grayscale* e matriz de intensidade respectiva a determinada área da mesma [18].

A resolução de uma imagem digital depende da dimensão da sua matriz, ou seja, com o aumento do número de pixels, a resolução aumenta. No entanto, o tratamento dessa matriz torna-se computacionalmente mais lento. Existem vários tipos de sensores capazes de obter imagem digital. Nas próximas subsecções serão abordados alguns desses tipos de sensores e explicado o seu funcionamento.

### 2.2.1 Sensores de imagem no espectro do visível

Para a captura de imagens digitais no espectro do visível são utilizados, principalmente, 2 tipos de sensores. O sensor *Charge-Coupled Device* (CCD) e *Complementary Metal-Oxide-Semiconductor* (CMOS). Cada um destes sensores é composto por milhões de transdutores fotossensíveis, cuja função é converter a energia luminosa em carga elétrica. Possuem ainda

uma superfície fotossensível, que recebe uma carga de luz permitindo captar a imagem, assim, quanto maior for a superfície fotossensível, melhor será a qualidade de imagem [19]. No entanto, estes sensores apenas conseguem medir a energia da radiação, sendo que, para que seja possível obter imagens a cores é necessário aplicar um filtro que permite direcionar cores específicas para pixels respetivos. O filtro mais comum é o filtro de *Bayer*. A Figura 2.5 mostra o funcionamento deste tipo de filtro.

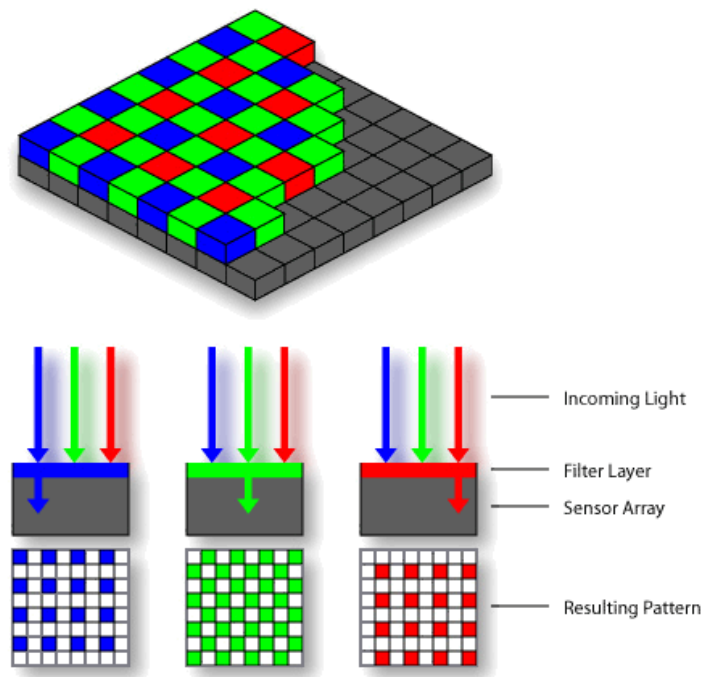


Figura 2.5: Aplicação de um filtro de *Bayer* de modo a obter uma imagem a cores [20].

O sensor CCD existe principalmente nas câmaras compactas, já o sensor CMOS está presente desde *webcams* simples e câmaras de *smartphones*, até às câmaras profissionais. A Figura 2.6 apresenta um exemplo de sensor CCD e CMOS.

## 2.2.2 Sensores especiais

Para além dos sensores falados anteriormente, existem ainda sensores especiais que permitem obter outra informação além da imagem a cores. Estes sensores são especialmente utilizados para o processamento de imagem em casos especiais como a deteção de distâncias e temperaturas. Ao longo desta secção serão explicados os modos de funcionamento desses



Figura 2.6: Exemplo de sensor CCD (esquerda) e CMOS (direita) [21].

sensores.

### **Térmicos**

Uma câmara térmica, ao contrário das câmaras no espectro do visível referidas anteriormente, são compostas por sensores capazes de capturar radiações no espectro do infravermelho, permitindo assim a criação de uma imagem de infravermelhos [22]. Normalmente é aplicada uma escala de cores a estas imagens, para que seja possível distinguir facilmente zonas quentes de frias. A Figura 2.7 mostra uma imagem térmica, obtida através de uma câmara térmica da marca *Flir* [23], com a respetiva escala de cores. Esta câmara é comercial, no entanto tem um elevado custo devido ao seu mercado específico e tecnologia utilizada no seu fabrico.

Este tipo de sensores podem ser utilizados mesmo em ambientes com pouca luminosidade, ao contrário de sensores como CCD e CMOS [22]. Existem assim diversas áreas onde estes se aplicam, desde a segurança, onde podem ser utilizados para detetar intrusos mesmo em situações de pouca iluminação [24], até à indústria, onde podem ser utilizados para detetar problemas de aquecimento em máquinas, que não são detetados pelo olho humano [22], passando pela deteção de pessoas através da temperatura do corpo humano [25].

### **Multi/Hiper Espectrais**

Os sensores Multiespectrais e Hiperespectrais são sensores que medem a energia produzida em várias bandas do espectro eletromagnético. A resolução espectral é o principal fator distintivo entre as imagens produzidas por estes dois tipos de sensores, sendo que os sensores



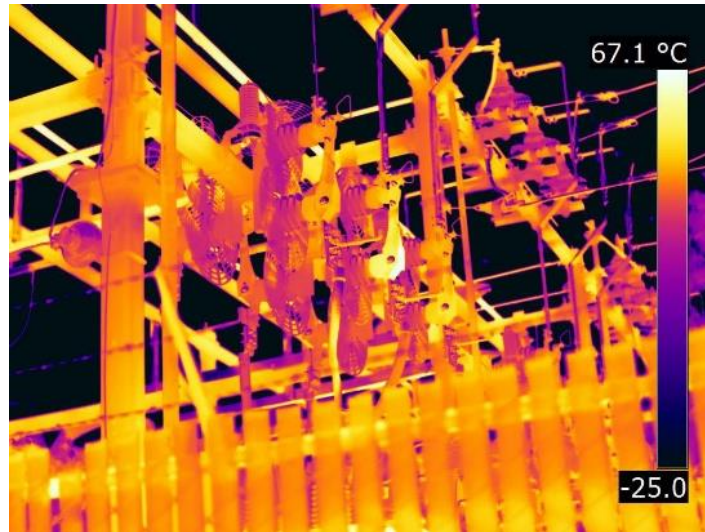


Figura 2.7: Imagem térmica com respectiva escala de cores, obtida através de uma câmara térmica da marca *Flir* [23].

hiperespectrais contêm um maior número de bandas com comprimentos de onda estreitos, fornecendo uma medição contínua em todo o espectro eletromagnético, enquanto que os sensores multiespectrais contêm, normalmente, entre 3 a 10 bandas, com amplos comprimentos de onda, em cada pixel da imagem produzida [26]. Deste modo, as imagens captadas por um sensor hiperspectral contêm mais dados do que as imagens captadas por sensores multiespectrais.

Num contexto prático, as imagens produzidas por sensores multiespectrais podem ser utilizadas, por exemplo, para mapear áreas florestais, enquanto que as imagens produzidas por sensores hiperespectrais podem ser usadas para mapear espécies de árvores dentro da mesma área florestal [27]. A Figura 2.8 mostra a comparação entre imagem multiespectral e hiperspectral.

## Distância

São diversos os tipos de sensores de imagem de distância. Estes tipos de sensores permitem obter imagens nas quais se percebem quais os objetos mais próximos e mais distantes, sendo possível representar as distâncias a cada objeto capturado pela câmara. Para isso existem 3 grandes tipos de sensores, os sensores chamados de *Time Of Flight* (TOF), *Structured Light* e *Stereo*.

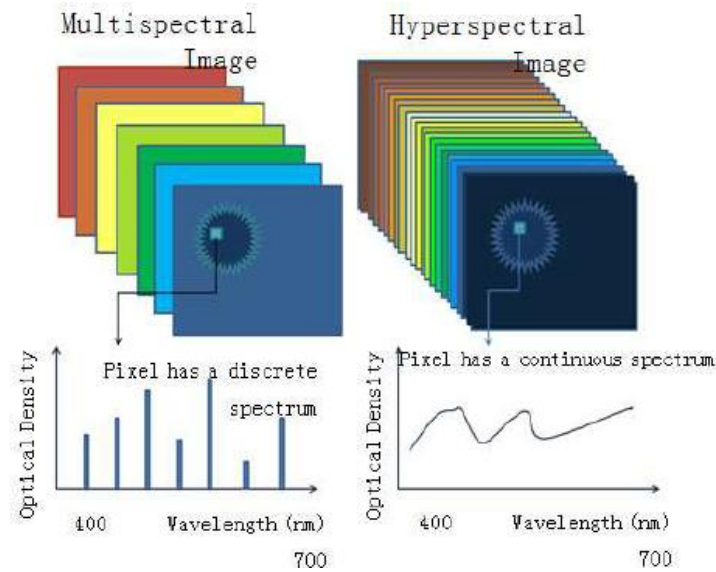


Figura 2.8: Comparação entre uma imagem multiespectral (esquerda) e hiperespectral (direita) [28].

Os sensores TOF funcionam sobre o princípio de envio e recepção de um sinal, medindo as propriedades do sinal recebido. Determinando o tempo de voo e, conseqüentemente, através deste tempo e da velocidade de propagação do sinal obtém-se a distância ao objeto [29].

Os sensores *Structured Light* funcionam através da projeção de um padrão numa cena, previamente estabelecido, permitindo ao sistema, através da captura desse mesmo padrão projetado, calcular a profundidade de cada pixel da imagem recebida. Este cálculo é efetuado através da deformação de cada ponto do padrão projetado em combinação com o padrão original [30].

Por fim, os sensores *Stereo* permitem obter imagens de distância através de duas lentes, a determinada distância, de modo a que as duas imagens capturadas possam ser processadas e comparadas, criando uma imagem 3D [31].

A Figura 2.9 apresenta um exemplo de câmara do tipo TOF (*SwissRange SR4000*), assim como uma imagem obtida por essa mesma câmara. Na Figura 2.10 é apresentado um exemplo de uma câmara do tipo *Structured Light* (*Kinect*) e um exemplo de imagem obtida por essa mesma câmara. A Figura 2.11 apresenta um exemplo de câmara do tipo *Stereo* (*Zed*) e um exemplo de imagem obtida pela mesma.

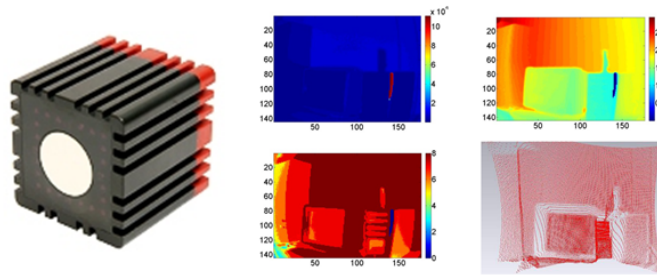


Figura 2.9: Câmara do tipo TOF (*SwissRange SR4000*) (esquerda) [32] e imagem obtida pela mesma (direita) [33].

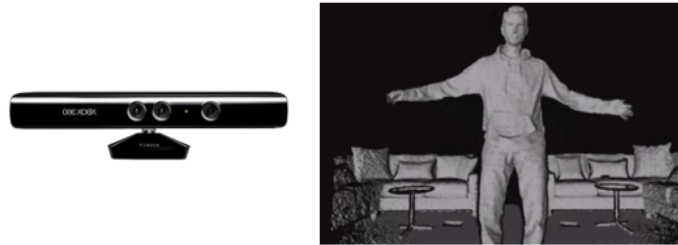


Figura 2.10: Câmara do tipo *Structured Light* (*Kinect*) (esquerda) [34] e imagem obtida pela mesma (direita) [35].



Figura 2.11: Câmara do tipo *Stereo* (*Zed*) (esquerda) e imagem obtida pela mesma (direita) [36].

## 2.3 Algoritmos para detecção de postura

Existe um elevado número de algoritmos de detecção da postura humana, no entanto poucos são aqueles que o fazem de forma dinâmica e em ambientes pouco controlados. Fala-se de ambiente controlado quando as posturas estão bem definidas, isto é, a pessoa sabe exatamente que posturas deverá assumir para testar o sistema e este só as detecta se forem bem representadas.

A utilização de imagem digital para detecção da postura tem sido cada vez mais explorada, uma vez que permite que tudo seja feito de uma forma não invasiva para o ser humano. Assim, não existindo dispositivos em contacto direto com o mesmo, as suas ações vão ser naturais, aumentando a precisão e credibilidade do algoritmo. Nos principais algoritmos de detecção de postura a detecção é feita recorrendo a objetos externos, como bandeiras [8], ou simplesmente através do ensinamento prévio do sistema para as posturas pretendidas [7].

Uma postura surge assim como o conjunto de localizações 2D ou 3D das articulações, sendo possível, através destas localizações, aferir a posição e deslocamento de todos os membros. No entanto, o problema que é comum a todos estes algoritmos prende-se com posições corporais críticas como: deitado, sentado, encolhido, de lado, etc. [7] [37] e em situações de grupos de pessoas, onde algumas partes do corpo se sobrepõem [37]. Neste tipo de posicionamento, a precisão destes sistemas baixa significativamente.

Todos os algoritmos de detecção de posturas apresentados têm como base vídeos ou um conjunto de imagens recolhidos a partir de câmaras digital. São assim vários os tipos de câmaras utilizadas entre estes algoritmos. Tal como descrito na secção anterior, estas câmaras podem se diferenciar no tipo de imagem que conseguem obter, no entanto, na atualidade, a *Kinect* é o dispositivo de eleição da maioria destes algoritmos, uma vez que o seu próprio Kit de Desenvolvimento de *Software* (SDK) é um dos mais utilizados no que diz respeito à detecção do esqueleto humano. A *Kinect* é constituída por uma câmara RGB, um sensor de profundidade, um acelerômetro de três eixos, um motor de inclinação e um vetor de microfones [38]. Assim é possível obter, com apenas um dispositivo, diferentes tipos de imagens. A Figura 2.12 mostra os diversos componentes de uma *Kinect*.

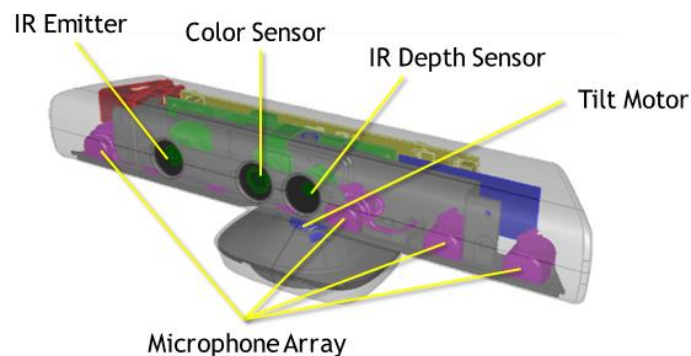


Figura 2.12: Configuração de *Hardware* de um dispositivo *Kinect* [39].

Tal como referido anteriormente, o *Software da Kinect, Kinect Skeletal Tracking* é bastante utilizado na deteção do esqueleto humano, sendo esta deteção realizada em três passos: No primeiro é feita uma análise, por pixel, de forma a detetar e classificar partes de corpo; Numa segunda fase é encontrado um centroide global de modo a definir as articulações do corpo; Por ultimo é feito um mapeamento das articulações, de forma a que se encaixem num esqueleto humano, através de dados previamente conhecidos acerca do esqueleto humano [40]. A Figura 2.13 ilustra os passos explicados anteriormente.

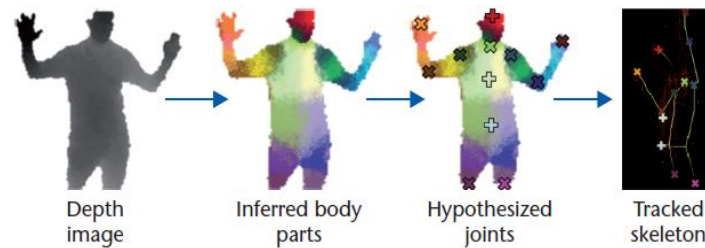


Figura 2.13: Passos de deteção do esqueleto humano através do *Software Kinect Skeletal Tracker* [40].

Em Abril de 2017, surgiu a biblioteca *OpenPose* [41]. Utilizando apenas imagem RGB, esta biblioteca é capaz de detetar e extrair valores 2D das partes principais do corpo humano. Nesta biblioteca, é possível efetuar uma deteção de corpo, cara e mãos, num total de 130 pontos de interesse possíveis, sendo 15 ou 18 deles para partes do corpo, 21 para cada mão e 70 para a cara.

Para a deteção corporal é utilizado um de dois conjuntos de dados (*Common Objects in Context* (COCO) ou *MPII Human Pose Dataset*) com imagens de pessoas, anotadas com o esqueleto humano, sendo ainda utilizado o *CMU Panoptic Dataset* [42] durante o desenvolvimento do algoritmo, visto que este contém cerca de 65 sequências de aproximadamente 5 horas e 30 minutos e 1.5 milhões de esqueletos 3D disponíveis. Esta deteção é feita através da abordagem descrita em [37], onde é utilizada uma rede neural que, recebendo uma imagem, faz em simultâneo uma previsão de mapas de confiança para deteção de partes do corpo (Figura 2.14b) e de campos de afinidade para associação das partes do corpo (Figura 2.14c), sendo que este processo é feito em várias etapas, para que essa deteção seja mais credível. De seguida, é executado um conjunto de combinações bipartidas para associação das partes do corpo, onde é utilizado um *score* que permite definir a que pessoa pertence a parte respetiva

e fazer uma conexão correta das partes em cada pessoa na imagem/frame (Figura 2.14d). Através desta abordagem é possível detetar várias pessoas na imagem e definir a sua postura. Por fim, com um algoritmo de inferência gananciosa são ligadas todas as partes e definidos os pontos 2D para cada uma das articulações (Figura 2.14e).

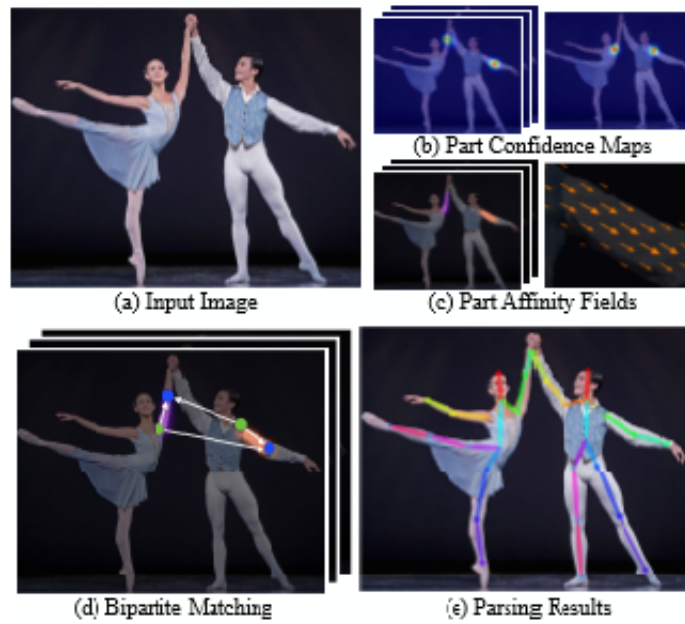


Figura 2.14: Detecção do esqueleto humano através da biblioteca *OpenPose* [37].

Em [43][44] são apresentadas abordagens de detecção de múltiplos esqueletos humanos em simples imagens RGB com resultados eficientes, no entanto ficam aquém de [37].

## 2.4 Instalação *OpenPose*

Devido ao elevado poder computacional requerido pela biblioteca apresentada, é necessário *Software* e *Hardware* específico. Relativamente ao *Hardware*, os requisitos mínimos são: placa gráfica *Nvidia* (CUDA) com pelo menos 1.6 Gb disponíveis, 2Gb de memória RAM e CPU com pelo menos 8 cores. Quanto ao *Software*, é necessária a biblioteca *OpenCV* [45], a *framework* de *Deep Learning Caffe* [46], a plataforma de programação paralela *CUDA* [47] e por fim a biblioteca de aceleração de GPU *CuDNN* [48]. A nível de Sistemas Operativos, podem ser utilizados: a distribuição de Linux *Ubuntu* (versão 14 ou 16), *Windows* (versão 8 ou 10) ou *Nvidia TX2*.

Para o desenvolvimento deste projeto de mestrado e de acordo com os requisitos da biblioteca *OpenPose*, foi construído um computador com uma placa gráfica *Nvidia 1080 Ti* (11 Gb), 32 Gb de memória RAM e um processador Intel core i7 com 8 cores. Relativamente ao *Software* foi utilizada a versão 3.3.1 do OpenCV, a versão *custom* do *Caffe*, *CUDA* versão 8.0 e *CuDNN* versão 5.1. O Sistema Operativo utilizado foi o Linux na distribuição *Ubuntu* 14.04. Este *Software* foi instalado e configurado recorrendo a um *playbook* criado para a plataforma de automação *Ansible* [49]. Este *playbook* permite, com apenas uma execução, configurar todo o sistema, sendo possível replicar a configuração para outras máquinas.





## Capítulo 3

# Cenário de Testes

Ao longo deste Capítulo serão apresentados todos os cenários de testes, assim como o *Software* desenvolvido para a aquisição de dados e sincronismo entre câmaras.

Todos os dados visuais utilizados durante o desenvolvimento deste projeto foram recolhidos no âmbito do estudo sobre “Conflito, emoções e eficácia em equipas de trabalho: um estudo experimental” aprovado pelo Conselho de Ética e Deontologia através do parecer 12/2018, a decorrer no Departamento de Educação e Psicologia da Universidade de Aveiro. Cada sessão envolve a participação de quatro voluntários, sentados em redor de uma mesa retangular, em posições previamente determinadas e estáticas, ou seja, ao longo de toda a sessão a posição de cada um dos voluntários é mantida.

De forma a ser possível analisar todos os participantes, foi necessário obter vídeos de vários ângulos distintos, optando-se pela instalação de três câmaras (*Kinect*) fixas para todas as sessões, cada uma ligada a um computador independente. Todos os computadores utilizados contêm a mesma configuração.

Para que se torne uniforme a análise dos resultados obtidos ao longo das sessões a estrutura da sala é mantida, marcando-se as posições de todos os elementos necessários (mesa, armários, etc.).

A Figura 3.1 representa a estrutura do laboratório onde foram realizadas as sessões.

Por fim, foram colocados dois leds em posições estratégicas da sala que servem para obter um *feedback* visual do início e fim de cada experiência.

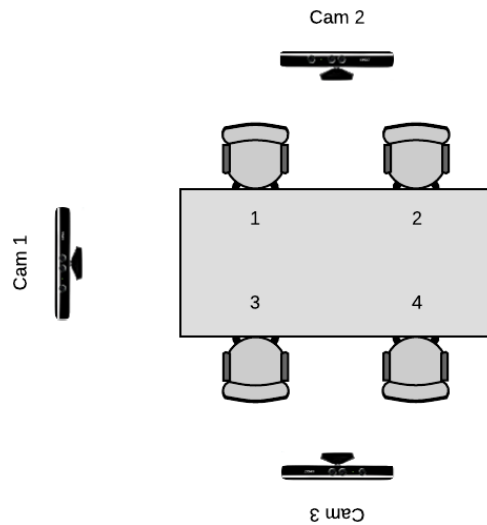


Figura 3.1: Organização do laboratório e esquema de posições dos id's estáticos.

---

### 3.1 Descrição do Protocolo

Tal como referido anteriormente, cada uma das experiências envolveu a participação de um grupo de quatro participantes (estudantes universitários), sendo dois deles do sexo masculino e os outros dois do sexo feminino. Aos participantes foi, inicialmente, dado a conhecer o consentimento informado, tendo todos eles consentido participar. Ainda no início da experiência, responderam a alguns questionários para despiste de doenças cardíacas e psicológicas, questões sócio-demográficas, escala de predisposição ao contágio, escala de emoções, escala de regulação emocional e questões sobre a experiência prévia com legos (dado ser essa a tarefa proposta, conforme indicado abaixo).

Depois de efetuados os procedimentos iniciais, seguiu-se a fase experimental. Esta fase foi constituída por duas etapas, que consistiu em duas tarefas de construção de legos que deveriam ser realizadas em grupo. A primeira tarefa incidiu sobre a construção de um cavalo, tendo os participantes 15 minutos para a sua realização. A segunda tarefa consistiu na construção de uma nave espacial, devendo os participantes começar por apresentar e discutir as ideias individuais relativas ao plano a seguir para, em conjunto, decidirem como proceder, dispondo de 5 minutos para o fazer. Após os 5 minutos de discussão, foram dados 20 minutos para

construção da nave espacial.

As experiências realizadas poderiam ser relativas a dois tipos de condições: a condição de controlo, onde não era induzido conflito e apenas se pretende avaliar de que forma a emergência de conflitos e as emoções associadas influenciam o funcionamento das equipas de trabalho; a condição de conflito de tarefa, onde, para a segunda tarefa foram dadas instruções dúbias e complexas, como a seguir se exemplifica: os participantes eram instruídos a distribuir tarefas pelos elementos do grupo; definir como deveria ser feita essa distribuição; era-lhes dito que o papel de cada elemento do grupo poderia variar ao longo da tarefa; que podiam construir mais do que um objeto, ou apenas um, mas cada um deveria ter de comprimento, aproximadamente, 30 cm; era-lhes indicado que a altura do(s) objeto(s) construído(s) é uma variável importante; que a nave espacial teria que ser aerodinâmica; que o tamanho do(s) motor(es) deveria permitir que a nave atingisse velocidades elevadas; que a nave espacial deveria ter cores quentes e deveria poder ser visualizada a partir da terra mas que, em voo, deveria ser invisível aos observadores.

Por fim, depois de realizadas todas as etapas, foi feito o *debriefing*, informando os participantes de acordo com o tipo de condição a que foram sujeitos.

### 3.2 *Software* de aquisição de dados

De forma a obter dados visuais das experiências realizadas, foi necessário desenvolver *Software* de aquisição de dados que permite obter vídeo das três câmaras em simultâneo.

Para que seja possível, após a recolha, efetuar uma sincronização dos vídeos, optou-se por se guardar a hora relativa a cada frame capturada na forma de *timestamp*, ou seja, com a data e hora sobre a forma de um valor inteiro. Deste modo, para que todos os computadores tenham a mesma hora, foi necessário configurar todos com o mesmo servidor de *Network Time Protocol* (NTP) para que o erro de sincronismo fosse mínimo. O NTP é um protocolo utilizado para manter a hora sempre correta, com uma elevada exatidão. Este protocolo é construído sobre o *Internet Protocol* (IP) e o *User Datagram Protocol* (UDP) [50].

Relativamente ao *Software* desenvolvido este foi dividido em duas partes. Inicialmente os dados foram adquiridos em *raw*, sem compressão nem processamento, ou seja, são guardados tal como captados pelo sensor da câmara. A captura dos dados da câmara utilizada foi feita através da biblioteca *OpenCV*, desenvolvida para fornecer uma infra-estrutura comum

para aplicações de visão computacional [45], com a *framework* *OpenNI*. Esta *framework* é *open-source* e fornece uma Interface de Programação de Aplicações (API) para a escrita de aplicações, sendo assim possível comunicar de forma simples com vários dispositivos. Neste projeto foi apenas utilizada para comunicar com a *Kinect* [51]. A aquisição, de forma a diminuir o espaço de armazenamento necessário, é feita de duas em duas frames (valor *default*), até que o utilizador pretenda terminar. A cada frame recolhida é guardado um *timestamp* num ficheiro à parte. É permitido ao utilizador que, através de argumentos, escolha o intervalo de gravação de frames, o número de frames a recolher e o nome do ficheiro de vídeo e de *timestamps* a gravar. Para que a gravação seja feita corretamente, é feito um *sleep* de 33 milissegundos entre frames.

A abordagem explicada anteriormente, apesar de gerar bons resultados, leva a que os ficheiros de vídeo ocupem demasiado espaço de armazenamento, sendo necessário uma enorme capacidade de armazenamento para que seja possível analisar todos os dados recolhidos. Assim, seguiu-se uma nova abordagem, onde os dados são guardados em formato *AVI*, utilizando, através do *OpenCV*, a biblioteca *FFMPEG* com o codec de vídeo *H.264*. Um codec de vídeo permite efetuar compressão e descompressão de vídeo de forma a economizar espaço, uma vez que este explora a redundância espacial do vídeo de forma a comprimir o mesmo da melhor forma possível. Desta forma, frames sucessivas cuja informação é idêntica são comprimidas numa pequena taxa de bits. O *H.264* utilizado pela biblioteca *FFMPEG* é do tipo *lossy*, ou seja, a compressão e descompressão do vídeo são feitas com possíveis perdas de informação, no entanto o resultado obtido é suficientemente parecido ao original [52]. Este passo é feito instanciando o objeto de *VideoWriter* com o *Four Character Code* (FOURCC) “H264” (Figura 3.2).

```
VideoWriter outputVideo;
// Declare name of file
const string filename = "VideoFile"+st+str+".avi";
//set video codec
int fourcc = CV_FOURCC('H','2','6','4');
int fps = 30;
outputVideo.open(filename, fourcc, fps, 5);
```

Figura 3.2: Instanciação do objeto *VideoWriter* com o codec *H.264*.

Nesta abordagem, são guardadas todas as frames, ao contrário do que era feito anterior-

mente, visto não ser necessário economizar espaço de armazenamento, uma vez que isso é feito através da compressão. Todo o restante programa funciona da mesma forma da abordagem anterior, guardando também todos os *timestamps* num ficheiro à parte.

A Figura 3.3 apresenta os diagramas de fluxo do *Software* de aquisição desenvolvido.

A Figura 3.4 representa o gráfico da média de armazenamento necessário por vídeo, utilizando cada uma das abordagens apresentadas. Para o cálculo desta média foram utilizados 21 vídeos.

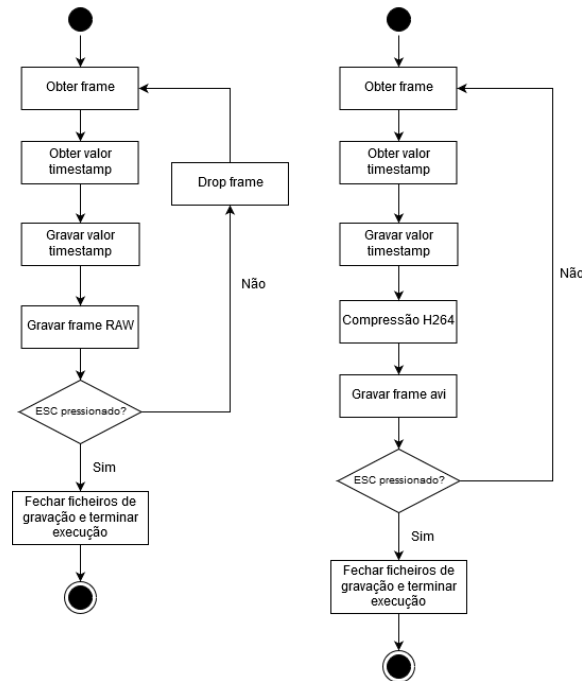


Figura 3.3: Diagrama de fluxo do *Software* de aquisição desenvolvido (na esquerda é representada a gravação do vídeo no formato *RAW* e na direita a gravação do vídeo comprimido no formato *AVI*).

Observando o gráfico é possível concluir que a compressão de vídeo permite reduzir o espaço necessário para armazenamento em cerca de 98.8%.

### 3.3 Sincronismo pós aquisição

Depois de recolhidos os dados há necessidade de sincronizar os 3 vídeos, visto que estes não se iniciam no mesmo instante e que, devido a condicionantes de captura (frames perdidas),

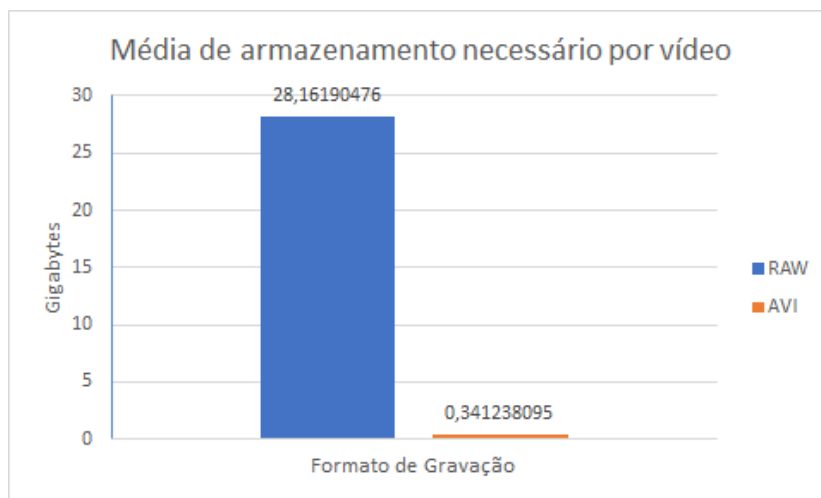


Figura 3.4: Gráfico da média de armazenamento necessário por vídeo em cada abordagem.

não mantêm o sincronismo ao longo do tempo. Estas condicionantes são, principalmente, a escrita de frames em disco, visto que este é um processo bloqueante e que, por vezes, leva à perda de uma ou outra frame estando bloqueado a escrever a frame anterior.

Os erros de escrita em disco poderiam ser diminuídos utilizando um disco *Solid State Drive* (SSD) em vez de um *Hard Disk Drive* (HDD). Um HDD é constituído por pratos metálicos que rodam a alta velocidade, possuindo uma cobertura magnética, onde a leitura e escrita é feita através de uma cabeça de leitura móvel. Já no SSD todo o sistema de pratos rotativos e cabeças de leitura é substituído por chips de memória flash [53], tornando o acesso e escrita de dados mais rápido que nos discos HDD, aumentando ainda a durabilidade do disco e diminuindo o ruído.

De forma a efetuar o sincronismo dos vídeos foi necessário desenvolver *Software* que permita avaliar para cada frame o respetivo *timestamp* e ir comparando, ao longo do tempo, os valores respetivos à frame em análise de cada câmara.

O *Software* desenvolvido permite visualizar, ao mesmo tempo, os vídeos recolhidos pelas três câmaras, tendo ainda acesso aos *timestamps* das frames que estão a ser apresentadas. Inicialmente, tendo acesso a todos os *timestamps* de cada câmara, é feito um sincronismo com o objetivo de alinhar todos os vídeos pelo que se inicia mais tarde, ou seja, são percorridos todos os ficheiros de *timestamps*, guardando aquele cujo primeiro valor é o mais alto e de seguida, olhando para os restantes ficheiros, é encontrada a frame cujo valor de *timestamp* difere

menos em relação ao valor guardado inicialmente. No final deste sincronismo o resultado que se pretende é um mapeamento entre número da câmara e frame inicial. Assim, a visualização de cada vídeo inicia-se na frame respetiva. Para manter o sincronismo ao longo do tempo, foi criado um critério de exclusão de frames, onde, em cada iteração, são comparados os três *timestamps* e determinado qual o valor mais alto. Depois disto, sempre que existir um valor que difere em 100 do valor mais alto, a frame associada não é mostrada nem guardada no resultado. O valor 100 surgiu após alguns testes efetuados no sentido de verificar qual o valor que melhor se adequava para que a sincronização fosse feita de forma não exaustiva mas precisa. Estes passos são feitos ao longo de todo o vídeo, comparando sempre os *timestamps* das três frames.

A Figura 3.5 representa o método de exclusão de frames explicado anteriormente.

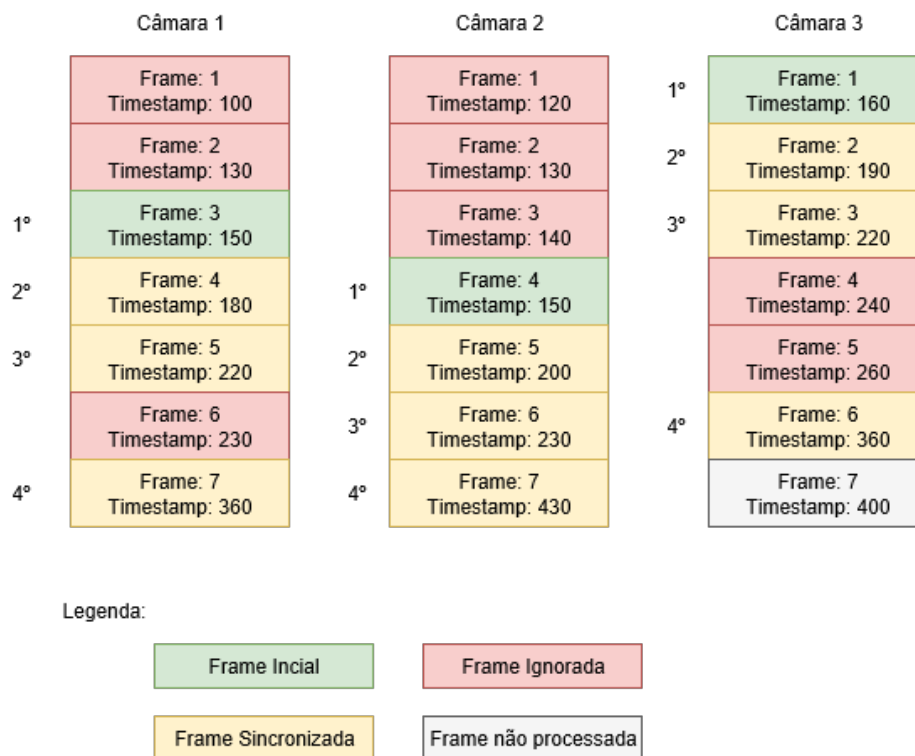


Figura 3.5: Método de exclusão de frames para efetuar um sincronismo automático dos vídeos.

Além de efetuar uma sincronização automática contínua, o *Software* desenvolvido tem opção de sincronização manual, fornecendo ao utilizador a hipótese de escolher um vídeo, com interação através do teclado, para avançar ou recuar de uma em uma frame. Com esta opção o utilizador tem assim a possibilidade de efetuar uma sincronização mais precisa.

São fornecidas também outras opções como o avanço e recuo de 10 em 10 ou de 100 em 100 frames em todos os vídeos, efetuar uma pausa nos mesmos e terminar a execução do *player*.

De forma a guardar o resultado da sincronização efetuada (frames e *timestamps*), removendo partes, iniciais e finais, não importantes dos vídeos, é fornecida a função de corte. Sempre que o utilizador pressiona a tecla respetiva pela primeira vez, é iniciado o corte do vídeo e o número de frames por segundo é aumentado, de modo a que os vídeos sejam processados mais rapidamente. Assim que o utilizador pretender terminar o corte e sincronização, deve apenas clicar novamente na tecla e o programa termina com a gravação dos vídeos cortados e sincronizados e dos respetivos *timestamps* em ficheiros distintos.

O *Software* desenvolvido aceita, como *input*, vídeos em formato *RAW* ou *AVI*, podendo o utilizador escolher através de argumentos. É ainda dada a opção do utilizador escolher um nome para adicionar aos ficheiros resultantes do corte explicado anteriormente.

O esquema de teclas de interação com o *Software* criado segue em Apêndice na Tabela A.1

### 3.4 Divisão de fases experimentais

Tal como referido na Secção 3.1, as experiências realizadas dividem-se em 2 partes. Assim, de modo a que seja possível efetuar uma comparação entre as 2 fases da experiência é necessário que seja feita uma separação nos vídeos para que essas 2 fases se distingam. Para isso, foi desenvolvido *Software* que permite, visualmente, definir qual a frame onde acaba a primeira fase da experiência e qual a frame onde começa a segunda fase da experiência, guardando os 2 valores num ficheiro de texto relativo à mesma. Este *Software* permite que o utilizador, ao ver o vídeo que inseriu como *input*, avance rapidamente no mesmo para que consiga ir direto às frames pretendidas, onde, clicando a primeira vez na tecla atribuída a esta função, é guardado o número dessa frame como final da primeira fase e, clicando novamente nessa tecla, é guardado o número da frame respetiva como início da segunda fase da experiência, terminando assim a execução do programa.

O esquema de teclas de interação com o *Software* criado segue em Apêndice na Tabela A.2



## Capítulo 4

# *Software* de Detecção

Neste Capítulo é explicado todo o *Software* de detecção desenvolvido. Iniciando-se pelo corte da região de interesse dos vídeos e obtenção dos pontos de interesse do corpo humano, até a extração das características pretendidas.

### 4.1 Região de Interesse

Com os vídeos sincronizados é agora possível efetuar uma análise detalhada de cada um. No entanto, antes de fazer o processamento utilizando a biblioteca *OpenPose* apresentada no Capítulo 2, é fundamental, para que o tempo de processamento seja reduzido, fazer um corte da região de interesse a analisar, uma vez que, como explicado anteriormente, as experiências decorrem apenas numa mesa que se mantém estática durante todas as recolhas, assim, todas as zonas em volta da mesa não são importantes para a análise. Desta forma foi desenvolvido *Software* que, através de ficheiros de configuração com informação acerca da região de interesse de cada câmara (Figura 4.1), fornecidos como *input* ao mesmo, juntamente com os vídeos que se pretende fazer o corte, fornece vídeos prontos a serem processados pelo *OpenPose* onde a região que não interessa está pintada a preto.

```
xmin 130  
xmax 520  
ymin 150  
ymax 470
```

Figura 4.1: Exemplo de configuração da região de interesse de uma câmara.

A Figura 4.2 mostra um exemplo do corte da região de interesse de uma frame obtida pela câmara 1.



Figura 4.2: Exemplo de corte da região de interesse de uma frame obtida pela câmara 1.

Na Figura 4.3 está representado o diagrama de fluxo do *Software* desenvolvido.

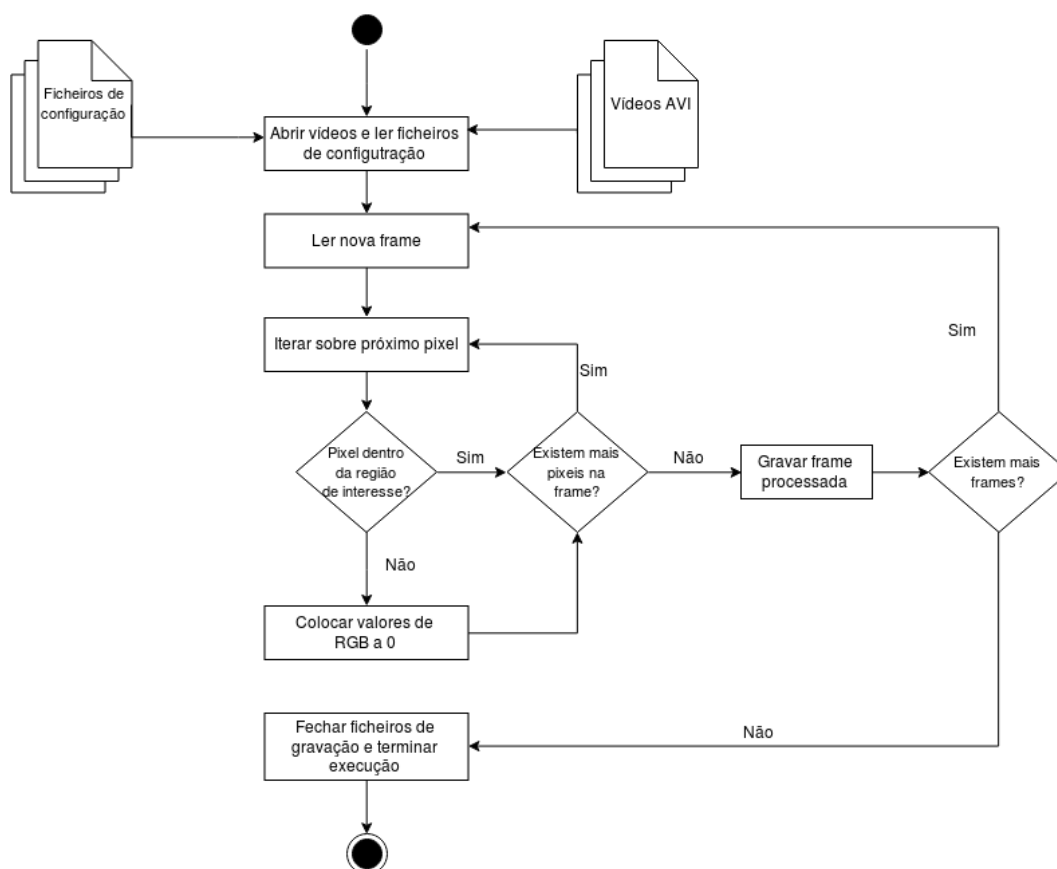


Figura 4.3: Diagrama de fluxo do *Software* de corte de região de interesse desenvolvido.

## 4.2 Pontos de Interesse

Depois de obtidos os vídeos com a região de interesse definida, segue-se a análise dos mesmos, com o objetivo de obter os pontos de interesse associados a cada pessoa presente em cada frame. Designa-se por ponto de interesse cada articulação do corpo humano, sendo detetadas pelo *OpenPose*, como explicado no Capítulo 2, um máximo de 18 articulações. No entanto, no contexto apresentado, apenas será analisada a parte superior do corpo a partir da cintura, uma vez que, durante todas as experiências, as pessoas se encontram sentadas. Como *output* do processamento resultam um conjunto de ficheiros *JavaScript Object Notation* (*JSON*) relativos aos pontos de interesse detetados em cada frame e um vídeo onde as posturas detetadas são apresentadas em cada frame do mesmo. Os ficheiros *JSON* são estruturados através de listas de valores com pares (chave/valor), tornando simples, rápido e eficiente o processamento das informações contidas nos mesmos. Cada ficheiro contém todos os pontos de interesse relativos às pessoas encontradas na frame respetiva, sobre a forma  $x, y, c, x, y, c, \dots$ , onde  $x$  e  $y$  representam as coordenadas de cada articulação e  $c$  representa a percentagem de confiança que o algoritmo tem sobre as coordenadas anteriores, pela ordem definida na biblioteca e que é mostrada na Figura 4.4. Na Figura 4.5 é apresentado um excerto de um dos ficheiros gerados pela biblioteca *OpenPose*, onde se pode verificar o resultado para a primeira pessoa detetada na frame respetiva a esse ficheiro.

## 4.3 Área Ocupada

Obtidos os ficheiros *JSON* é necessário proceder agora à sua análise, de forma a perceber a posição de cada pessoa, sendo assim possível obter as características pretendidas. Deste modo, foi desenvolvido *Software* de processamento que tem como *input* os vídeos obtidos e já pré-processados das três câmaras, os ficheiros *JSON* associados a cada vídeo e os ficheiros de configuração de cada câmara. Nos ficheiros de configuração, para além das coordenadas da região de interesse, estão descritas as posições das pessoas por id estático, ou seja, para a pessoa cujo id é 1 está descrita a sua posição sobre a forma de, por exemplo,  $x_{min} y_{max}$ , de modo a que seja possível, posteriormente, fazer uma atribuição de id's estáticos, conseguindo-se assim manter os mesmos id's para cada pessoa nas três câmaras. A Figura 4.6 representa o ficheiro de configuração da câmara 2. Na Figura 3.1 é apresentado o esquema de posições

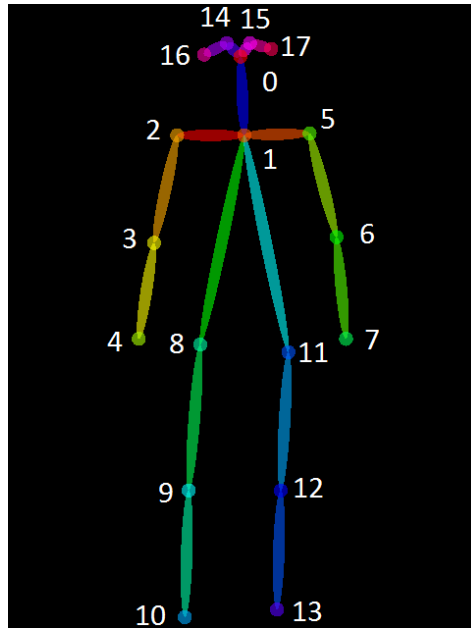


Figura 4.4: Imagem representativa da ordem dos pontos de interesse definidos pela biblioteca *OpenPose* [10].

```
{
  "version":1.0,
  "people":[
    {
      "pose_keypoints":[
        428.587,205.398,0.868436,
        444.239,219.774,0.866439,
        415.569,221.048,0.85868,
        410.345,256.31,0.88299,
        402.514,285.016,0.809752,
        472.992,218.448,0.769775,
        483.402,258.94,0.427595,
        499.069,275.946,0.103947,
        432.533,295.47,0.276504,
        0,0,0,
        0,0,0,
        463.816,298.106,0.256644,
        0,0,0,
        0,0,0,
        423.385,200.186,0.901855,
        433.834,201.468,0.910918,
        0,0,0,
        453.338,198.905,0.797334
      ],
      1,
    }
  ]
}
```

Figura 4.5: Excerto de um ficheiro de pontos de interesse *JSON* gerado pela biblioteca *OpenPose*.

de id's estáticos criado.

Olhando para a figura descrita anteriormente é possível observar que a pessoa com o id 1, na câmara 2 é vista na posição  $x$  máximo e  $y$  máximo. Assim, na primeira linha, respetiva

```
xmin 80
xmax 540
ymin 110
ymax 470
xmax ymax
xmin ymax
xmax ymin
xmin ymin|
```

Figura 4.6: Ficheiro de configuração da câmara 2 (região de interesse e mapeamento dos id's estáticos).

---

ao mapeamento dos id's, do ficheiro de configuração é colocado “*xmax ymax*”.

Este *Software* começa por abrir os ficheiros de configuração das 3 câmaras, de forma a obter as posições explicadas anteriormente para que sejam utilizadas posteriormente. De seguida, são percorridos todos os ficheiros *JSON*, sendo estes processados para as 3 câmaras em simultâneo, ou seja, são abertos, em paralelo, os 3 ficheiros relativos às 3 frames em análise. Por exemplo, iniciando-se na frame 0 é observada esta frame em cada um dos 3 vídeos, assim como os respetivos ficheiros *JSON*. Nesta análise, começa-se por identificar o número de pessoas detetadas em cada uma das frames, verificando-se se o número de pessoas é maior que 1 e menor que 5, caso contrário as frames são descartadas, passando-se para a iteração seguinte. Caso o número de pessoas detetadas se encontre dentro dos valores definidos, então, para cada pessoa, é atribuído um id provisório e são guardadas as coordenadas dos 14 pontos de interesse em estudo, obtendo-se ainda o valor de  $x$  e  $y$  máximos da pessoa em análise. Com estes valores mínimos e máximos são determinadas a largura, altura e área ocupada pela pessoa, guardando-se esses mesmos valores num ficheiro de *output* da frame e câmara respetiva. Este ficheiro contém ainda os valores dos pontos de interesse,  $x$  e  $y$  mínimo e máximo referidos acima.

Depois de serem processados todos os valores das pessoas reconhecidas na frame em análise, e utilizando esses mesmos valores, são desenhados retângulos (*bounding boxes*), em volta de cada pessoa, de modo a que se consiga gerar um vídeo de *output* que permita verificar, visualmente, todas as deteções efetuadas ao longo do vídeo. Num passo seguinte, utilizando os valores de configuração obtidos inicialmente, onde, para cada câmara, se definem as posições

de cada id estático, é feito um mapeamento entre os id's provisórios associados a cada pessoa detetada com os respetivos id's estáticos, dependendo da localização de cada um. Para a atribuição de id's é feita uma divisão da região de interesse em 4 partes, determinando-se qual a pessoa que está dentro de cada uma das partes. Assim, com esta informação e a informação de qual o id a atribuir em cada parte da imagem, torna-se direto o mapeamento referido anteriormente, mesmo quando são detetadas apenas 2 ou 3 pessoas. A Figura 4.7 mostra a divisão efetuada e a respetiva atribuição de id's da câmara 2. Estes id's são ainda colocados na frame de *output*, centrados na *bounding box* desenhada anteriormente e o mapeamento é guardado no ficheiro de *output*, juntamente com as coordenadas referidas inicialmente.



Figura 4.7: Divisão e atribuição de id's estáticos relativos à câmara 2.

Assim, como *output* deste *Software*, surge um conjunto de ficheiros de texto (1 por cada conjunto de 3 frames relativas a cada uma das câmaras) e um conjunto de frames onde são representados as *bounding boxes* em volta de cada pessoa detetada com o respetivo id estático.

## 4.4 Sobreposições

Depois de identificadas as pessoas são estudadas as sobreposições entre as mesmas, de forma a aferir uma possível interação entre os elementos do grupo. A deteção de sobreposições envolve a estimativa de um valor de profundidade, visto que todos os valores obtidos são 2D. Assim, é feito um ciclo que permite efetuar uma verificação entre todas as combinações possíveis de pessoas, duas a duas. A estimativa é feita com base nos valores mínimos e máximos de  $x$  e  $y$  das duas pessoas a comparar, assim como os valores de  $x$  e  $y$  dos membros

superiores de cada uma.

Para utilização posterior, é verificado inicialmente se a abertura dos membros superiores é idêntica nas duas pessoas a comparar. De seguida é verificado se existe sobreposição entre as duas *bounding boxes*. Se não existir então é assumido que não existe interação e a execução desta função termina. Caso exista, assume-se uma possível interação, incrementa-se a percentagem de certeza de interação em 30%, valor este que diz respeito a um aumento de certeza notável, no entanto não garante uma sobreposição, uma vez que ainda não foi feita uma estimativa de profundidade, e a função continua a sua execução verificando se a distância entre os valores de  $y$  máximos das duas *bounding boxes* é relativamente pequena. Se não, então é assumido que as duas pessoas não estão no mesmo plano em  $z$  e a execução da função termina com uma não sobreposição/interação. Se sim, então o nível de confiança em relação à sobreposição aumenta 25%, tendo-se assim mais certeza na sobreposição, existindo já uma estimativa de profundidade, e é verificado de seguida, caso as aberturas dos membros sejam idênticas (que aumenta a confiança em 5%, valor este que é baixo devido a este não ser um elemento fundamental para a estimativa de sobreposições), se a área das duas *bounding boxes* é idêntica. Caso não seja, então a execução termina com uma possível interação de confiança baixa (60%). Caso seja, aumenta-se a confiança em 25%, visto que se tem ainda mais certeza de que as pessoas pertencem ao mesmo plano em  $z$ , e é verificada se a distância em  $y$  é ainda mais próxima, aumentando 10% o nível de confiança. Desta forma a execução da função termina com uma confiança muito elevada (85% ou 95%) de que as pessoas estão sobrepostas ou em interação. A Figura 4.8 representa, graficamente, a deteção de uma sobreposição.

A Figura 4.9 representa o diagrama de *workflow* da deteção de sobreposições explicada.

O resultado desta deteção é guardado no final de cada ficheiro de *output* descrito anteriormente, sendo ainda criados 3 *heatmaps* (um por cada câmara), que permitem, visualmente, perceber as regiões onde existiu um maior número de interações/sobreposições. Designa-se por *heatmap* uma representação gráfica de dados, onde os valores são representados por uma escala de cores. A criação do *heatmap* que permite verificar as zonas com mais interações é feita através de uma matriz iniciada a zeros e com a dimensão igual às imagens em estudo. Sempre que haja uma possível sobreposição, são incrementados todos os valores referentes à região sobreposta na matriz referida. No final todos os valores são convertidos numa escala de 0 a 255 onde se utiliza um *colorMap* que permite distinguir zonas com mais interações

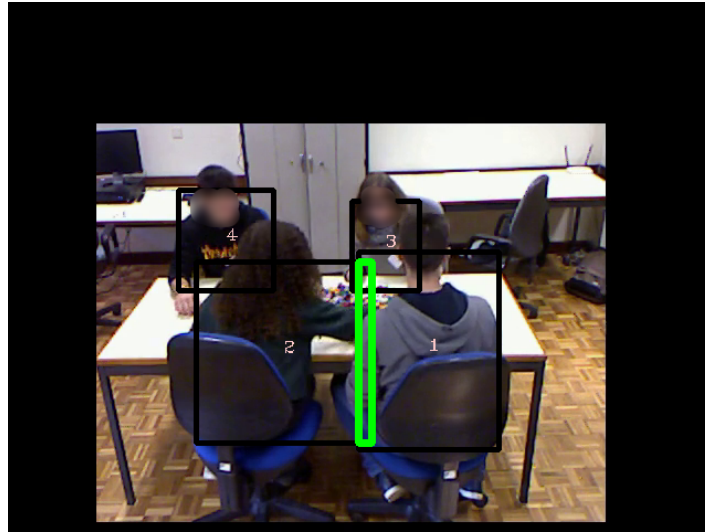


Figura 4.8: Representação gráfica de uma sobreposição.

---

(vermelho) de zonas com menos interações (azul). A Figura 4.10 representa a escala de cores utilizada.

## 4.5 Deslocamento

De forma a determinar o deslocamento das partes do corpo ao longo de cada fase da experiência foi desenvolvido *Software* que, recebendo como *input* os ficheiros gerados pelo *Software* apresentado na Secção 4.3 e os ficheiros que contêm o número das frames que dividem as duas fases da experiência, calcula o deslocamento de cada parte de forma individual. As partes do corpo (vetores) são definidas através de um par de pontos de interesse que representam duas articulações do corpo ligadas, num total de 13 vetores, tal como representado na Figura 4.11. O cálculo é feito frame a frame, analisando a frame atual e a anterior. Desta forma é possível, através do cálculo da distância euclidiana, determinar o deslocamento, relativamente ao tempo, de cada vetor. No entanto, existe o problema da falha de deteção de alguns pontos de interesse em algumas das frames. Caso este problema aconteça, então é guardado o valor da última deteção do ponto de interesse em causa e feito o cálculo assim que este for novamente detetado. Desta forma, o *Software* permite ainda calcular a velocidade de deslocamento, dividindo a distância pelo valor do intervalo de frames em que o vetor não foi encontrado.



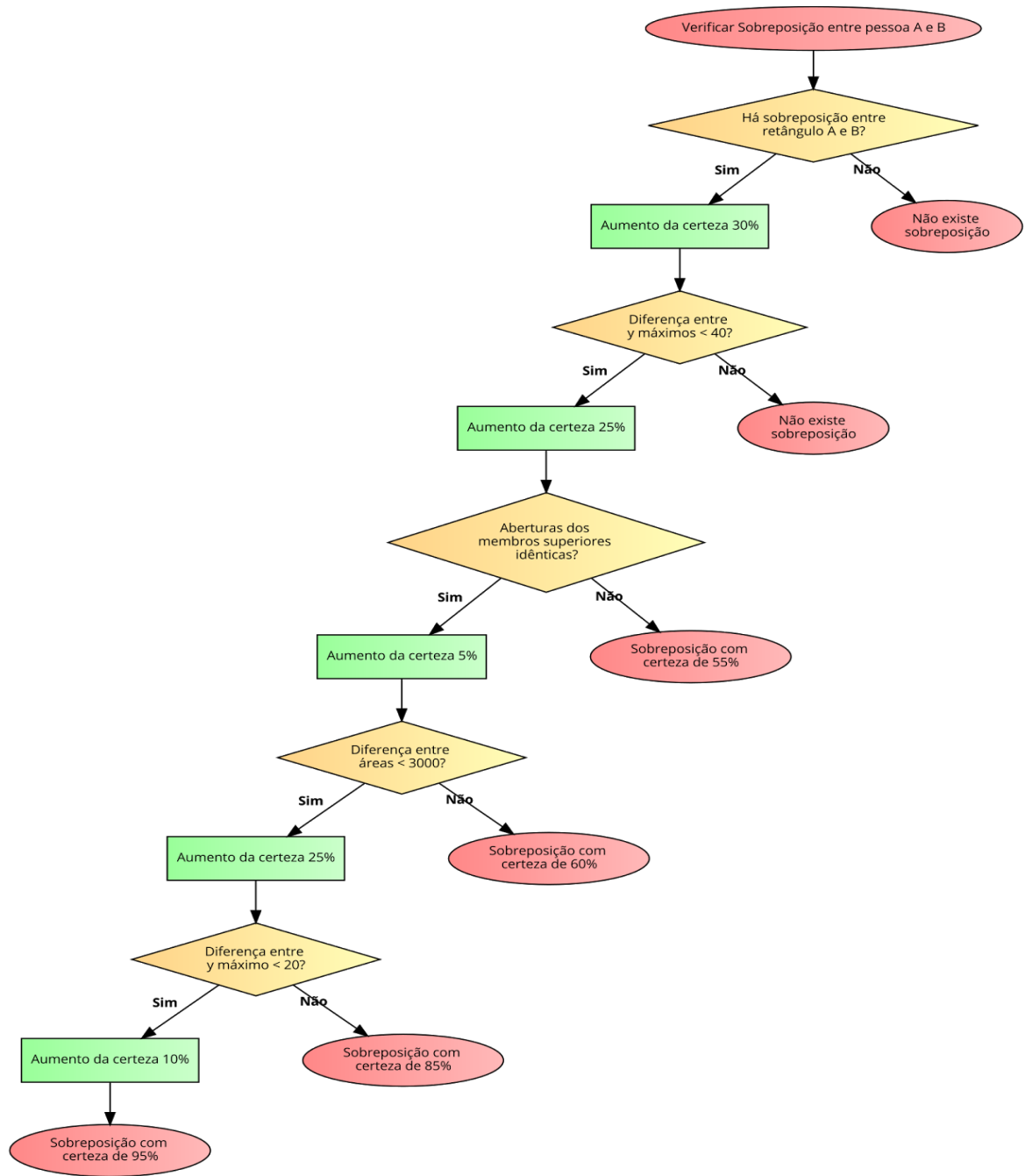


Figura 4.9: Diagrama de *workflow* da detecção de sobreposições.



Figura 4.10: *ColorMap* utilizado - *ColorMap-Jet* [54]

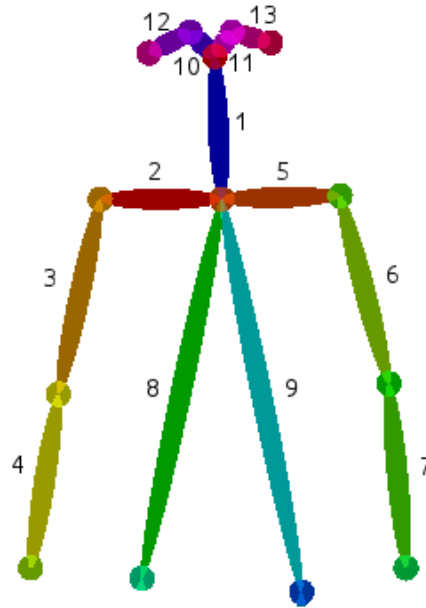


Figura 4.11: Representação dos vetores associados às 13 partes do corpo em estudo.

Os valores relativos à distância são guardados em ficheiros de texto (um para cada câmara e para cada pessoa) respeitando a ordem crescente do número do vetor, sendo por fim escrito o número da frame e a área ocupada pela pessoa. Caso não seja possível efetuar o cálculo de algum dos vetores na frame em questão, então é colocado um '-' na sua posição. A Figura 4.12 mostra a estrutura dos ficheiros gravados.

```
v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 número_frame área
2.05083 0.948099 0.946084 1.96875 0.658313 0.647253 0.644313 - - 2.05487 2.3243 - 1.5438 1 2435.19
1.30623 0.670196 3.25587 3.53079 0.651619 0.664788 3.25489 - - 1.4376 1.30085 - 0.654445 2 2391.77
4.69665 11.473 - - 6.4784 9.74326 10.2919 - - 3.44898 4.22409 - 3.36505 3 1915.57
0.0449817 0.0303524 - - 1.29186 2.73602 1.95413 - - 1.29449 1.29329 - 1.29008 4 1955.12
1.50968 1.52128 2.90011 3.51573 1.53033 2.77267 3.14591 - - 1.2467 0.651362 - 0.0621075 5 2041.53
2.23749 6.91137 6.91714 2.14217 2.92535 3.27761 2.58604 - - 0.906458 0.90267 - 1.53795 6 1923.88
1.91853 0.661248 - - 1.93603 3.78028 3.23807 - - 2.06597 2.87966 - 3.47834 7 1844.62
3.22705 3.879 - - 2.58833 2.614 3.1425 - - 1.27241 0.664837 - 1.30349 8 1799.93
1.43369 1.45527 - - 2.05618 1.29169 0.0409688 - - 1.27634 0.66186 - 0.656629 9 1800.72
3.99941 4.41142 22.8909 - 5.76166 3.56186 1.49997 - - 0.911257 0.643713 - 1.26849 10 2120.53
1.26572 1.58481 18.9733 19.025 1.57887 1.85125 1.57965 - - 0.605432 0.605612 - 1.27168 11 1904.29
1.89766 1.91382 1.91325 1.47376 1.2961 3.9051 4.53053 - - 0.633797 1.25718 - 1.27416 12 2020.32
```

Figura 4.12: Estrutura do ficheiro de deslocamentos e áreas relativas à câmara 2 (pessoa com id 3).

Na Figura 4.13 está representado o processo de obtenção do gráfico dos deslocamentos, em relação ao tempo, durante toda a experiência, do vetor 4 (assinalado a vermelho nas frames demonstrativas) da pessoa com id 2 na câmara 2, sendo que, à esquerda do gráfico se encontra a

nuvem de pontos relativa à primeira fase e à direita a nuvem de pontos relativa à segunda fase. De forma a demonstrar este processo são apresentadas 3 frames que representam, visualmente, o deslocamento do vetor em estudo. Este gráfico é obtido através da ferramenta de criação de gráficos através da linha de comandos em Linux, *GnuPlot* [55].

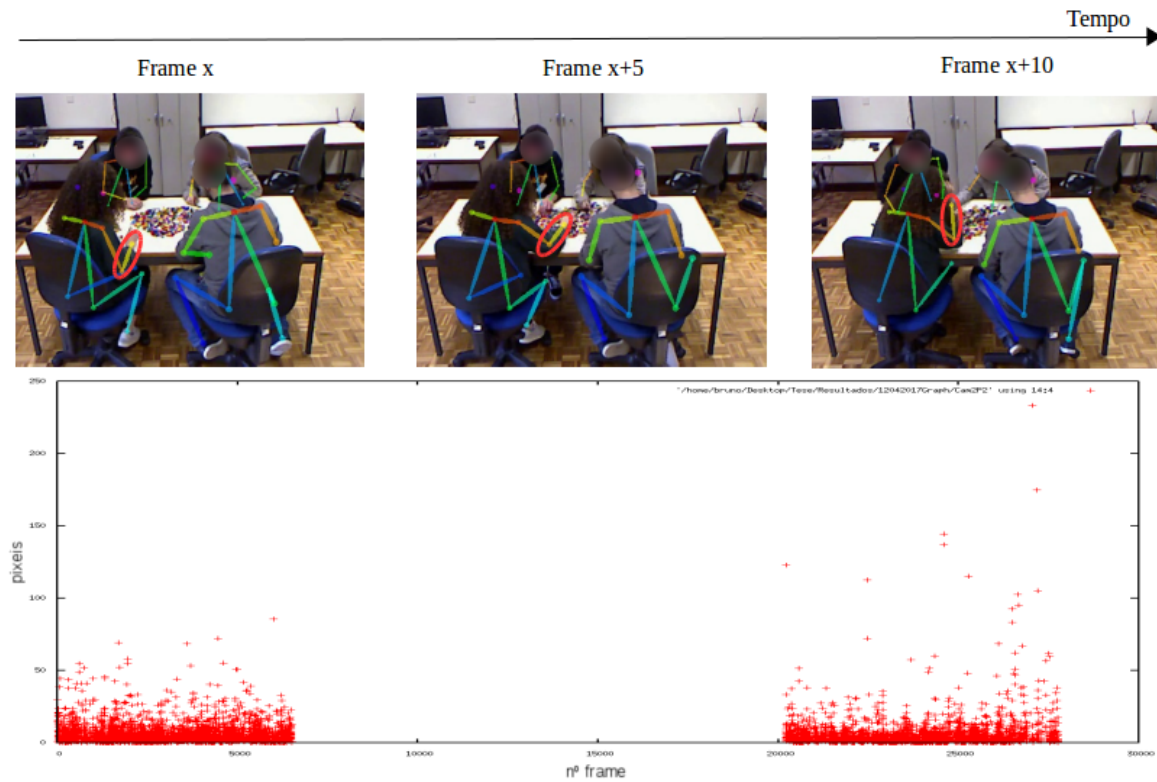


Figura 4.13: Processo de obtenção do gráfico dos deslocamentos para o vetor 4 da pessoa com id 2 na câmara 2.



## Capítulo 5

# Resultados

Neste Capítulo serão apresentados os resultados obtidos a nível de deslocamento dos vetores, área ocupada e sobreposições encontradas ao longo das experiências. Estes resultados dizem respeito a duas experiências, numa delas foi induzido conflito de tarefa e na outra não foi induzido qualquer tipo de conflito. Existem ainda mais resultados relativos aos dois tipos referidos, que seguem em Apêndice a este documento (Apêndice B), onde os resultados, apesar de terem algumas variações, são, em média, semelhantes aos apresentados de seguida.

### 5.1 Deslocamento

De forma a verificar a evolução dos deslocamentos, foram criados gráficos que demonstram os valores dos mesmos em cada frame. Estes valores podem ser relativos a cada vetor ou à média de todos os vetores encontrados. Para obter este tipo de resultados foi desenvolvido *Software* que permite ao utilizador escolher a experiência, câmara, pessoa e vetor que pretende. De seguida é gerado o gráfico respetivo através do *Software Gnuplot* [55]. Este gráfico encontra-se dividido em duas partes, onde a primeira parte se refere à primeira fase da experiência e a segunda parte referente à segunda fase da experiência. A Figura 5.1 representa a divisão efetuada. Pretende-se assim que, através da visualização dos gráficos, se conclua em que fase da experiência existiu um maior movimento por parte de cada participante.

Tal como referido no Capítulo 3 existem 3 câmaras no laboratório, sendo assim possível obter vídeos de vários ângulos diferentes. Desta forma, uma vez que são conhecidas as posições das pessoas, através dos id's estáticos e câmaras, ao efetuar a análise desses vídeos terá de se ter

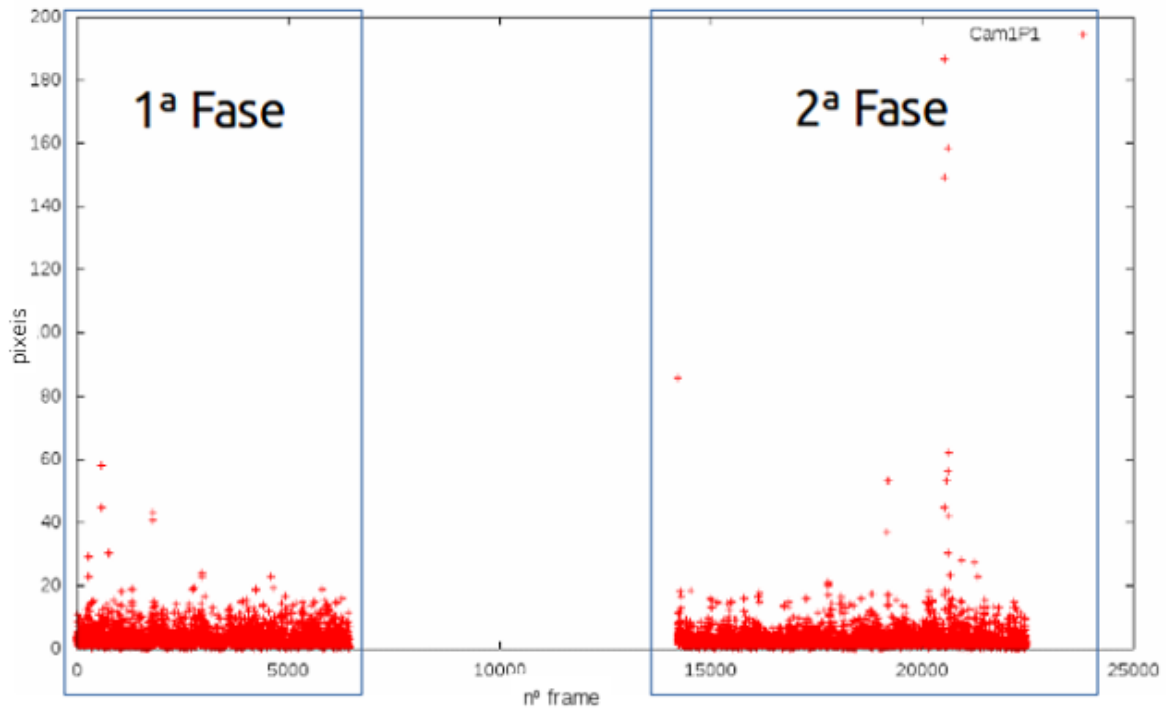


Figura 5.1: Divisão dos gráficos obtidos pelas duas fases da experiência.

em conta a posição de cada câmara em relação à pessoa a analisar, ou seja, ao analisar o vídeo da câmara 1, os resultados irão ser melhores para as pessoas com os id's 1 e 3 (Figura 3.1). Desta forma, a análise deve ser feita tendo em conta estas mesmas posições, obtendo-se assim resultados mais credíveis. As Figuras 5.2, 5.3, 5.4 e 5.5 mostram os gráficos relativos ao movimento dos participantes durante uma experiência em que não foi induzido qualquer tipo de conflito. As Figuras 5.6, 5.7, 5.8 e 5.9 mostram os gráficos relativos ao movimento dos participantes durante uma experiência em que foi induzido o conflito de tarefa. Através da observação destes gráficos verifica-se que, apesar de pequena, há uma diferença maior de movimento entre as duas fases da experiência quando há indução de conflito.

## 5.2 Área Ocupada

De forma a observar a variação de área ocupada por cada participante foram também criados gráficos com os dados obtidos. Assim, é possível verificar se determinado participante alterou a sua postura entre as duas fases da experiência, ocupando mais ou menos espaço. As Figuras 5.10, 5.11, 5.12 e 5.13 mostram os gráficos relativos à área dos participantes durante

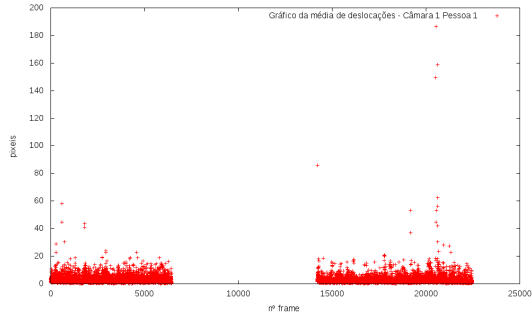


Figura 5.2: Gráfico relativo à média de deslocamentos da pessoa com id 1 numa experiência sem indução de conflito (13/12/2017).

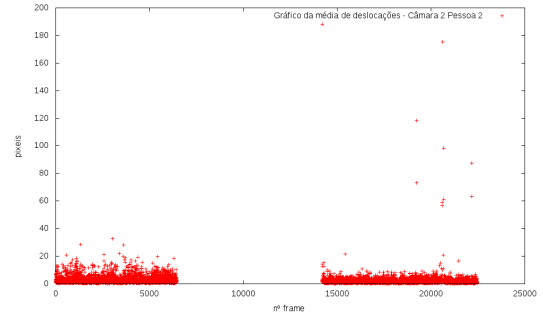


Figura 5.3: Gráfico relativo à média de deslocamentos da pessoa com id 2 numa experiência sem indução de conflito (13/12/2017).

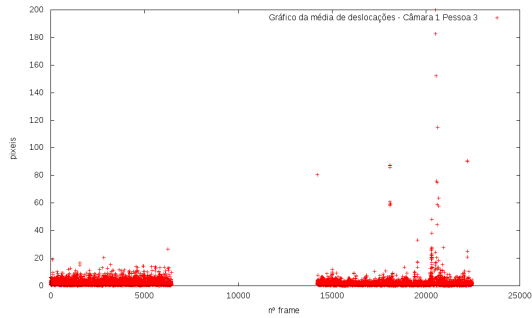


Figura 5.4: Gráfico relativo à média de deslocamentos da pessoa com id 3 numa experiência sem indução de conflito (13/12/2017).

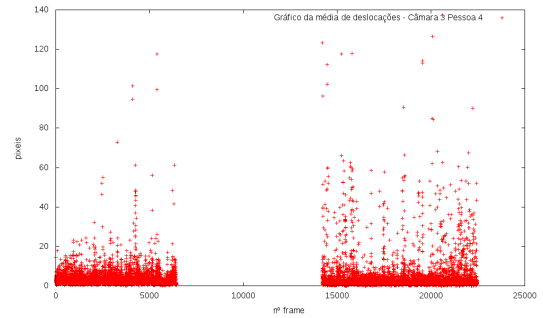


Figura 5.5: Gráfico relativo à média de deslocamentos da pessoa com id 4 numa experiência sem indução de conflito (13/12/2017).

---

uma experiência em que não foi induzido qualquer tipo de conflito. As Figuras 5.14, 5.15, 5.16 e 5.17 mostram os gráficos relativos à área dos participantes durante uma experiência em que foi induzido o conflito de tarefa. Novamente, através da observação destes gráficos, verifica-se que a variação da área entre as duas fases da experiência é, em média, maior quando há uma indução de conflito.

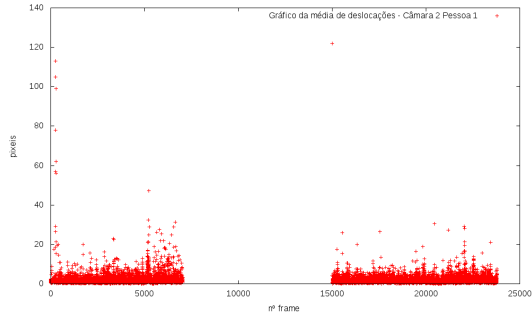


Figura 5.6: Gráfico relativo à média de deslocamentos da pessoa com id 1 numa experiência com indução de conflito de tarefa (15/12/2017).

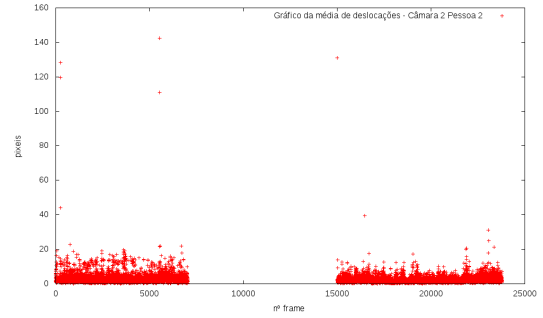


Figura 5.7: Gráfico relativo à média de deslocamentos da pessoa com id 2 numa experiência com indução de conflito de tarefa (15/12/2017).

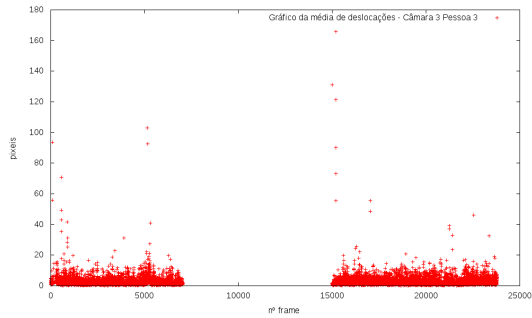


Figura 5.8: Gráfico relativo à média de deslocamentos da pessoa com id 3 numa experiência com indução de conflito de tarefa (15/12/2017).

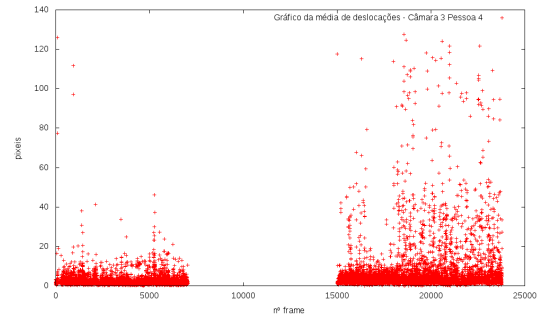


Figura 5.9: Gráfico relativo à média de deslocamentos da pessoa com id 4 numa experiência com indução de conflito de tarefa (15/12/2017).

### 5.3 Interação

Por fim, de forma a que seja possível observar o nível de interação entre os participantes, foi criado um *Heatmap* por câmara em cada experiência. Assim, é possível verificar quais os participantes com maior interação, podendo-se concluir quais foram os que tiveram um papel mais importante durante o decorrer da mesma. É necessário, durante a análise, ter novamente em conta a posição dos participantes relativamente à câmara. Observando cada *Heatmap* verifica-se, rapidamente, as zonas com maior interação, ainda que surjam alguns erros mínimos relativos a falhas na identificação das pessoas durante o processo de deteção.



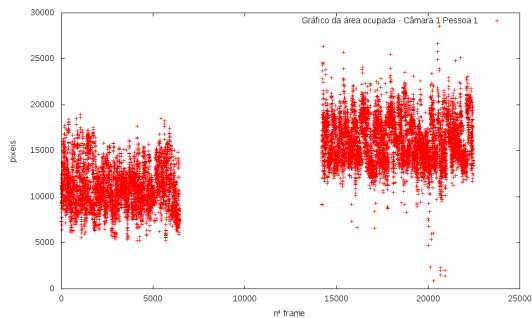


Figura 5.10: Gráfico relativo à área ocupada pela pessoa com id 1 numa experiência sem indução de conflito (13/12/2017).

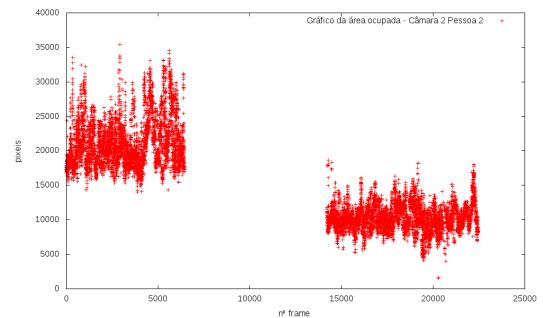


Figura 5.11: Gráfico relativo à área ocupada pela pessoa com id 2 numa experiência sem indução de conflito (13/12/2017).

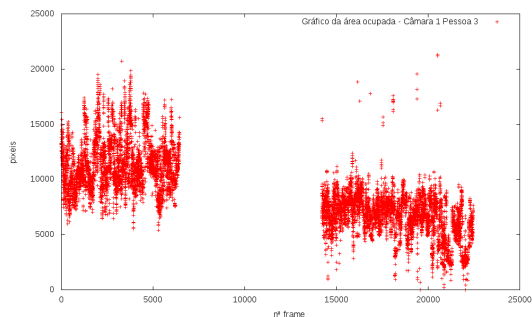


Figura 5.12: Gráfico relativo à área ocupada pela pessoa com id 3 numa experiência sem indução de conflito (13/12/2017).

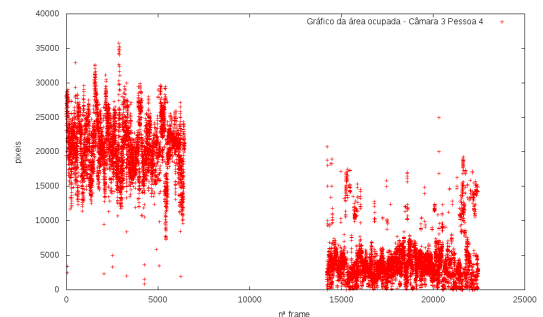


Figura 5.13: Gráfico relativo à área ocupada pela pessoa com id 4 numa experiência sem indução de conflito (13/12/2017).

---

Por exemplo, ao olhar para as Figuras 5.18, 5.19 e 5.20 verifica-se que as pessoas 1 e 2 foram as que mais interagiram. Já as pessoas 3 e 4 tiveram um número de interações bastante inferior. Comparando as Figuras 5.18, 5.19 e 5.20, que dizem respeito a uma experiência onde não foi induzido conflito, com as Figuras 5.21, 5.22 e 5.23, que dizem respeito a uma experiência onde foi induzido conflito de tarefa, percebe-se que o número de interações/sobreposições é maior quando há indução de conflito de tarefa.

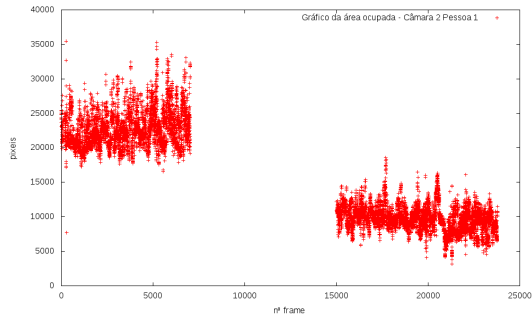


Figura 5.14: Gráfico relativo à área ocupada pela pessoa com id 1 numa experiência com indução de conflito de tarefa (15/12/2017).

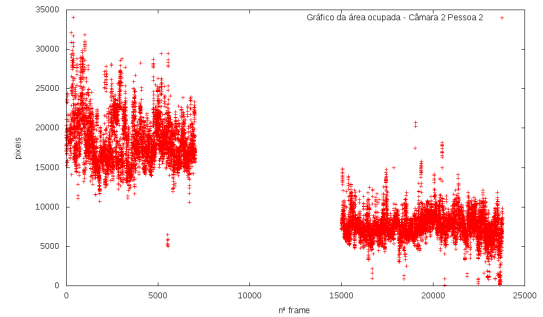


Figura 5.15: Gráfico relativo à área ocupada pela pessoa com id 2 numa experiência com indução de conflito de tarefa (15/12/2017).

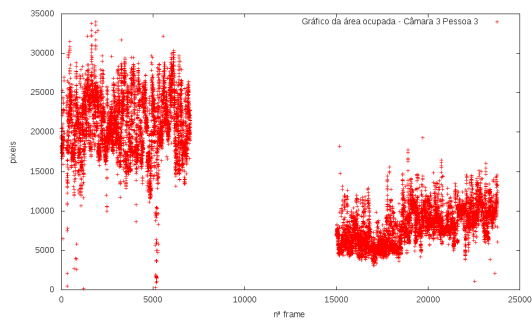


Figura 5.16: Gráfico relativo à área ocupada pela pessoa com id 3 numa experiência com indução de conflito de tarefa (15/12/2017).

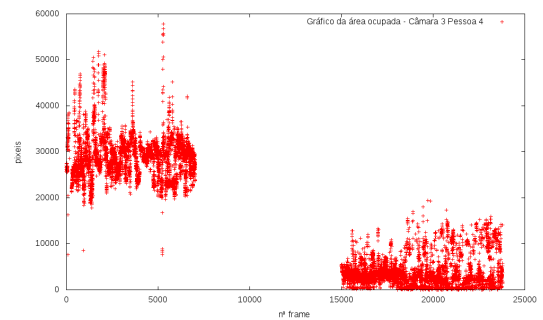


Figura 5.17: Gráfico relativo à área ocupada pela pessoa com id 4 numa experiência com indução de conflito de tarefa (15/12/2017).



Figura 5.18: *Heatmap* relativo à câmara 1 numa experiência sem indução de conflito (13/12/2017).

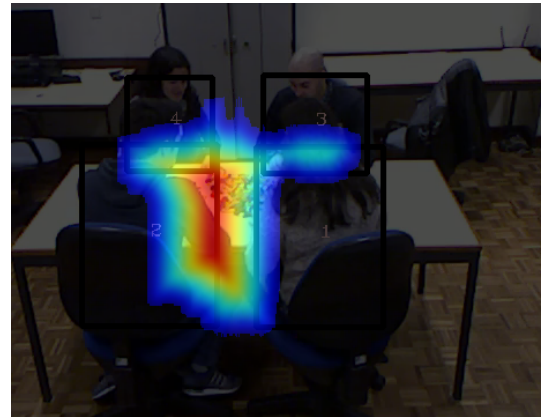


Figura 5.19: *Heatmap* relativo à câmara 2 numa experiência sem indução de conflito (13/12/2017).

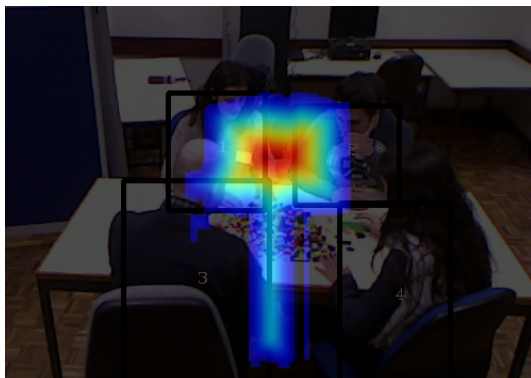


Figura 5.20: *Heatmap* relativo à câmara 3 numa experiência sem indução de conflito (13/12/2017).



Figura 5.21: *Heatmap* relativo à câmara 1 numa experiência com indução de conflito de tarefa (15/12/2017).

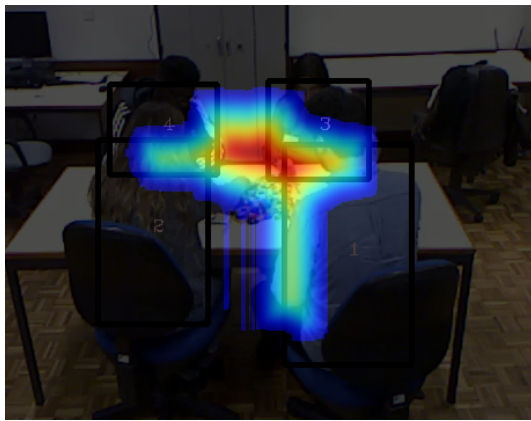


Figura 5.22: *Heatmap* relativo à câmara 2 numa experiência com indução de conflito de tarefa (15/12/2017).

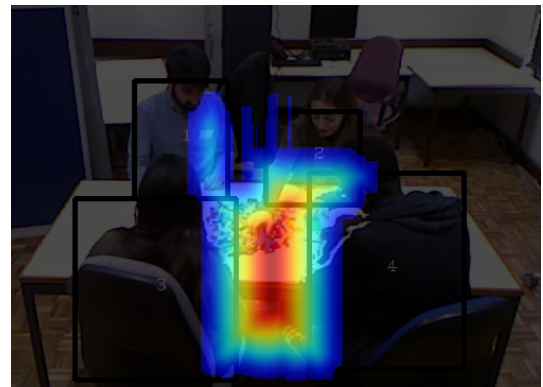


Figura 5.23: *Heatmap* relativo à câmara 3 numa experiência com indução de conflito de tarefa (15/12/2017).

## Capítulo 6

# Conclusão

O desenvolvimento deste projeto de mestrado permitiu obter uma resposta à pergunta “Como utilizar um PC e uma câmara para medir a postura corporal do corpo humano?”. Através da biblioteca *OpenPose* foi possível obter, de forma rápida, informações acerca de todas as articulações do corpo humano. Sendo esta uma biblioteca que permite, com apenas imagens ou vídeos em duas dimensões, detetar o número de pessoas presentes e obter a posição de cada ponto chave das pessoas encontradas, revolucionou de uma forma drástica os algoritmos de deteção do esqueleto humano. Devido ao enorme e anotado *dataset* de treino, a biblioteca apresenta resultados muito satisfatórios, mesmo em situações complicadas, onde as pessoas se apresentam em posições críticas de deteção, como: sentadas, deitadas, sobrepostas, etc. Desta forma, depois de analisados os resultados que a mesma poderia fornecer, foi decidido a sua utilização, que em muito ajudou no resultado final do sistema apresentado.

A obtenção de dados através da experiência a decorrer no Departamento de Educação e Psicologia foi feita de forma a que fosse possível alimentar o *OpenPose* com esses mesmos dados, sendo que, inicialmente, e sem estar definida a utilização desta biblioteca, foi decidido efetuar a captura de imagens em formato *RAW*, utilizando os dispositivos *Kinect*, sendo possível iniciar a recolha desses dados em paralelo com o processo de escolha da biblioteca, permitindo assim que os dados já recolhidos fossem utilizados para qualquer que fosse a escolha. Depois de escolhida a biblioteca foi decidido efetuar a aquisição em formato *AVI*, visto que este formato pode ser diretamente utilizado como *input*, permitindo ainda economizar espaço de armazenamento.

Alimentando o *OpenPose* com os dados recolhidos, foi possível obter um *output* satisfa-

tório, tanto em termos visuais com as frames anotadas com o esqueleto reconhecido, como em termos de localização das articulações de cada pessoa, individualmente, através de ficheiros *JSON*. Esses ficheiros permitiram alimentar o *Software* desenvolvido, sendo possível obter informações diretas como o número de pessoas detetadas na frame em questão, como informações indiretas, por exemplo, a área ocupada por determinado participante. Para o contexto apresentado não é necessário obter outro tipo de *output*, mesmo que a biblioteca permita recolher, por exemplo, apenas imagens dos esqueletos detetados ou ficheiro em outros tipos de formato. Até este ponto, o trabalho de aquisição de dados e deteção do esqueleto humano está completo.

Analisando, visualmente, o *output* referido anteriormente, é perceptível que o *OpenPose* consegue lidar com situações de conflito e sobreposição entre pessoas, tal como se desejava inicialmente, conseguindo ainda estimar a posição de partes ocultas de cada participante, sendo estes dados importantes, uma vez que as pessoas envolvidas se encontram sentadas e em constante interação. Assim, sempre que há uma sobreposição entre partes do corpo de duas ou mais pessoas, é possível, tecnologicamente, reconhecer e avaliar essas situações.

Com todos os dados recolhidos conseguiu-se extrair características que permitem aferir as posturas e emoções associadas, ou seja, olhando para um gráfico relativo à área ocupada por uma pessoa durante uma experiência pode-se concluir se a mesma assumiu uma postura mais contraída ou expansiva. Olhando ainda para um gráfico relativo ao movimento das articulações do corpo é possível de deduzir se a pessoa se encontra mais tímida ou mais confiante. Por fim, percebe-se, através da observação dos *heatmaps* criados, que o nível de interações entre os participantes poderá estar diretamente relacionado com a quantidade de movimento e a postura assumida por cada um, ou seja, se o indivíduo A se movimentou pouco, então o mesmo irá ter poucas interações. Por outro lado, se dois indivíduos B e C, que estão lado a lado, se movimentaram muito, então as interações nessa zona vão ser elevadas.

Olhando para os resultados obtidos no final deste projeto pode-se responder superficialmente à segunda pergunta colocada: “Como classificar cada postura como estando associada a determinada emoção?”, percebendo-se que há uma relação direta entre as posturas e as emoções associadas. Essa relação é perceptível através das características extraídas, onde, por exemplo, se uma pessoa se movimenta pouco, então o seu número de interações vai ser menor e assim pode-se assumir que a pessoa se encontra mais tímida, sendo possível fazer uso

destas características para vários tipos de análise. Visto que esta é uma área muito pouco explorada, este trabalho conseguiu contribuir, de uma forma bastante positiva para aquilo que é o avanço das investigações na mesma. Partindo pela recolha de informações sobre os sistemas existentes, passando pelos algoritmos de deteção do esqueleto humano e a escolha do algoritmo que mais se adequou ao que seria necessário, até à criação do sistema de extração das características.

As expectativas iniciais desta dissertação foram atingidas na sua generalidade, com a criação de um sistema que permite obter as características necessárias para a avaliação da postura humana, completando ainda com um caso de estudo em que foi possível o tratamento dos resultados da postura obtidos e a sua análise de acordo com o comportamento dos participantes. Os dois artigos publicados, [9] e [5], no âmbito deste tema, permitiram aprofundar ainda mais conhecimentos nesta área. Mesmo [5] não estando diretamente relacionado com o contexto de trabalho em grupo aqui apresentado, permitiu perceber que realmente se poderiam avaliar as emoções expressadas pelo corpo. Futuramente poderá ser adicionada a extração de outro tipo de características, que permitam analisar mais detalhadamente todas as posturas. Poderão ainda se recolher, simultaneamente, outro tipo de parâmetros do comportamento humano, por exemplo, colocação de voz, que permitam consolidar a ideia de que a postura está diretamente relacionada às emoções.





# Bibliografia

- [1] B. de Gelder, A. W. de Borst, and R. Watson. The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(2):149–158, 2014.
- [2] Anthony P Atkinsonô, Winand H Dittrichô, Andrew J Gemmell, and Andrew W Young. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33:717–746, 2004.
- [3] Winand H Dittrich, Tom Troscianko, Stephen E G Lea, and Dawn Morgan. Perception of Emotion from Dynamic Point-Light Displays Represented in Dance. *Perception*, 25(6):727–738, jun 1996.
- [4] A Rui Gomes, Maria João Silva, Salomé Mourisco, Susana Silva, Alfredo Mota, and Nuno Montenegro. Problemas e desafios no exercício da actividade docente: Um estudo sobre o stresse, "burnout", saúde física e satisfação profissional em professores do 3º ciclo e ensino secundário. *Revista Portuguesa de Educação*, 19(1):67–93, 2006.
- [5] Rui Sacchetti, Tiago Teixeira, Bruno Barbosa, António J R Neves, Sandra C Soares, and Isabel D Dimas. Human Body Posture Detection in Context : The Case of Teaching and Learning Environments. (SIGNAL 2018, The Third International Conference on Advances in Signal, Image and Video Processing):79–84, 2018.
- [6] Thi-Lan Lan Le, Minh-Quoc Quoc Nguyen, and Thi-Thanh-Mai Thanh Mai Nguyen. Human posture recognition using human skeleton provided by Kinect. *2013 International Conference on Computing, Management and Telecommunications (ComManTel)*, pages 340–345, 2013.

- [7] Zequn Zhang, Yuanning Liu, Ao Li, and Minghui Wang. A Novel Method for User-Defined Human Posture Recognition Using Kinect. *International Congress on Image and Signal Processing*, pages 736–740, 2014.
- [8] Chia Wen Chang, Min Da Nian, Yi Fang Chen, Ching Ho Chi, and Chin Wang Tao. Design of a Kinect Sensor Based Posture Recognition System. *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 856–859, 2014.
- [9] Bruno Barbosa, António J R Neves, Sandra C Soares, and Isabel D Dimas. Analysis of Emotions from Body Postures Based on Digital Imaging. (SIGNAL 2018, The Third International Conference on Advances in Signal, Image and Video Processing):73–78, 2018.
- [10] GitHub - CMU-Perceptual-Computing-Lab/openpose: OpenPose: Real-time multi-person keypoint detection library for body, face, and hands estimation. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>. [Online; Accessed 2018-05-17].
- [11] Mirko Raca and Pierre Dillenbourg. Classroom Social Signal Analysis. *Journal of Learning Analytics*, 1(3):176–178, 2014.
- [12] Christian Thureau and Vaclav Hlavac. Pose primitive based human action recognition in videos or still images. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, jun 2008.
- [13] Tao Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208–1221, sep 2004.
- [14] Rc Gonzalez and Re Woods. Digital image processing and computer vision. *Computer Vision, Graphics, and Image Processing*, 49(1):122, jan 1990.
- [15] Maria Lyra, Agapi Ploussi, and Antonios Georgantzoglou. MATLAB as a Tool in Nuclear Medicine Image Processing. *MATLAB - A Ubiquitous Tool for the Practical Engineer*, (October 2011):477–500, 2011.

- [16] Free How to Photoshop Tutorials, Videos & Lessons to learn Photoshop training | Photoshop Course. <http://www.we-r-here.com/ps/tutorials/>. [Online; Accessed 2017-11-28].
- [17] Greg Borenstein. *Making Things See: 3D Vision with Kinect, Processing, Arduino, and MakerBot*. 2012.
- [18] Naushadsblog: Image Processing and Computer vision in java (Image Filtering part1). <https://naushadsblog.wordpress.com/>. [Online; Accessed 2017-11-28].
- [19] Nicolas Blanc. CCD versus CMOS - has CCD imaging come to an end? *Photogrammetric Week 2001*, pages 131–137, 2001.
- [20] Wikimedia Commons. <https://commons.wikimedia.org/wiki/Main{ }Page>. [Online; Accessed 2017-11-28].
- [21] Photography tips and tricks, Equipment, Photography News, Photography Books, Tutorial, and Lighting - OneSlidePhotography.com. <http://oneslidephotography.com/>. [Online; Accessed 2017-11-28].
- [22] Wai Kit Wong, Poi Ngee Tan, Chu Kiong Loo, and Way Soong Lim. An effective surveillance system using thermal camera. *2009 International Conference on Signal Acquisition and Processing, ICSAP 2009*, pages 13–17, 2009.
- [23] FLIR Systems | Thermal Imaging, Night Vision and Infrared Camera Systems. <http://www.flir.com/home/>. [Online; Accessed 2018-02-06].
- [24] Tomasz Sosnowski, Grzegorz Bieszczad, and Henryk Madura. Image Processing in Thermal Cameras. pages 35–57. Springer, Cham, 2018.
- [25] Soonmin Hwang, Jaesik Park, Namil Kim, Yookyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June:1037–1045, 2015.
- [26] Lisa-Jen Ferrato and K. Wayne Forsythe. Comparing Hyperspectral and Multispectral Imagery for Land Classification of the Lower Don River, Toronto. *Journal of Geography and Geology*, 5(1):92–107, 2013.

- [27] What is the difference between multispectral and hyperspectral imagery? - eXtension. <http://articles.extension.org/pages/40073/what-is-the-difference-between-multispectral-and-hyperspectral-imagery>. [Online; Accessed 2017-11-28].
- [28] Moataz Aboras, Hani Amasha, and Issa Ibraheem. Early detection of melanoma using multispectral imaging and artificial intelligence techniques Early detection of melanoma using multispectral imaging and artificial intelligence techniques. *Http://Www.Sciencepublishinggroup.Com*, 3(November 2016):29, 2015.
- [29] S. Burak Gokturk, Hakan Yalcin, and Cyrus Bamji. A time-of-flight depth sensor - System description, issues and solutions. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2004-Janua(January), 2004.
- [30] Pietro Zanuttigh, Carlo Dal Mutto, Ludovico Minto, Giulio Marin, Fabio Dominio, and Guido Maria Cortelazzo. *Time-of-flight and structured light depth cameras: Technology and applications*. 2016.
- [31] G. Calin and V. O. Roda. Real-time disparity map extraction in a dual head stereo vision system. *Latin American Applied Research*, 37(1):21–24, 2007.
- [32] ITCS (IT Consulting and Services) Mesa Imaging SwissRange SR4000 - ITCS (IT Consulting and Services). <https://www.itcs.com.pk/product/mesa-imaging-swissrange-sr4000/>. [Online; Accessed 2018-06-13].
- [33] Dario Piatti and Fulvio Rinaudo. SR-4000 and CamCube3.0 Time of Flight (ToF) Cameras: Tests and Comparison. *Remote Sensing*, 4(4):1069–1089, apr 2012.
- [34] Kinect Sensor. <https://msdn.microsoft.com/en-us/library/hh438998.aspx>. [Online; Accessed 2018-05-16].
- [35] Xbox One Kinect camera can see through clothes and detect every part of anatomy | Daily Mail Online. <http://www.dailymail.co.uk/sciencetech/article-2503474/Xbox-One-Kinect-camera-clothes-detect-anatomy.html>. [Online; Accessed 2018-06-13].

- [36] ZED stereo camera simulates human visual depth perception. <https://thenextweb.com/dd/2015/05/19/zed-stereo-camera-simulates-human-visual-depth-perception/>. [Online; Accessed 2018-06-13].
- [37] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1302–1310. IEEE, jul 2017.
- [38] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *IEEE Transactions on Cybernetics*, 43(5):1318–1334, oct 2013.
- [39] Kinect for Windows Sensor Components and Specifications. <https://msdn.microsoft.com/en-us/library/jj131033.aspx>. [Online; Accessed 2017-12-02].
- [40] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE Multimedia*, 19(2):4–10, 2012.
- [41] OpenPose - Realtime Multiperson 2D Keypoint Detection from Video | Flintbox. <https://cmu.flintbox.com/public/project/47343/>. [Online; Accessed 2017-12-03].
- [42] CMU Panoptic Dataset. <http://domedb.perception.cs.cmu.edu/>. [Online; Accessed 2018-07-08].
- [43] Eldar Insafutdinov, Mykhaylo Andriluka, Leonid Pishchulin, Siyu Tang, Evgeny Levinkov, Bjoern Andres, and Bernt Schiele. ArtTrack: Articulated Multi-person Tracking in the Wild. dec 2016.
- [44] Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model. may 2016.
- [45] OpenCV library. <https://opencv.org/>. [Online; Accessed 2018-05-03].
- [46] Caffe | Deep Learning Framework. <http://caffe.berkeleyvision.org/>. [Online; Accessed 2018-05-29].

- [47] CUDA Zone | NVIDIA Developer. <https://developer.nvidia.com/cuda-zone>, 2011. [Online; Accessed 2018-05-29].
- [48] NVIDIA cuDNN | NVIDIA Developer. <https://developer.nvidia.com/cudnn>. [Online; Accessed 2018-05-29].
- [49] Ansible is Simple IT Automation. <https://www.ansible.com/>. [Online; Accessed 2018-05-29].
- [50] D.L. Mills. Internet time synchronization: the network time protocol. *IEEE Transactions on Communications*, 39(10):1482–1493, 1991.
- [51] OpenNI on your PC | Kinect Hacks. <http://www.kinecthacks.com/guides/install-openni-on-your-pc/>. [Online; Accessed 2018-04-30].
- [52] OpenCV: cv::VideoWriter Class Reference. [https://docs.opencv.org/3.4.1/dd/d9e/classcv\\_1\\_1VideoWriter.html](https://docs.opencv.org/3.4.1/dd/d9e/classcv_1_1VideoWriter.html). [Online; Accessed 2018-06-19].
- [53] SSD vs HDD – Diferenças e vantagens e desvantagens de cada um. – PCManias.com. <http://www.pcmánias.com/ssd-vs-hdd-diferencas-e-vantagens-e-desvantagens-de-cada-um/>. [Online; Accessed 2018-05-04].
- [54] Rainbow Colors Indigo – images free download - Help Your User Retain Your Message | HOW Design. <http://fwfx.info/rainbow-colors-indigo.html>. [Online; Accessed 2018-05-15].
- [55] Philipp K. Janert. *Gnuplot in Action: Understanding Data With Graphs*. 2009.

## Apêndice A: Teclas de interação do *Software* desenvolvido.

### A.1 *Software* de sincronismo

Tabela A.1: Teclas de interação do *Software* de sincronismo.

Tecla(s)	Funcionalidade
'1', '2' ou '3'	Selecionar vídeo para avançar ou recuar de uma em uma frame
'e' ou 'q'	Avançar ou recuar de uma em uma frame, respetivamente
'd' ou 'a'	Avançar ou recuar de 10 em 10 frames, respetivamente
'x' ou 'z'	Avançar ou recuar de 100 em 100 frames, respetivamente
'p'	Play/Pausa
'q'	Terminar execução do <i>Software</i>
'c'	Corte dos vídeos (1º clique inicia o corte e 2º clique termina o corte)

### A.2 *Software* de divisão das fases experimentais

Tabela A.2: Teclas de interação do *Software* de divisão das fases experimentais.

Tecla	Funcionalidade
'c'	Guardar número da frame atual (1º clique guarda final da 1ª fase e 2º clique guarda início da 2ª fase)

## Apêndice B: Gráficos e *Heatmaps* obtidos relativamente às diferentes experiências realizadas.

### B.1 Média de deslocamentos em experiências sem indução de conflito

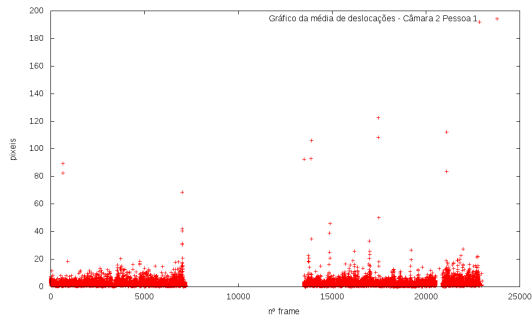


Figura B.1: Gráfico relativo à média de deslocamentos da pessoa com id 1 numa experiência sem indução de conflito (20/12/2017).

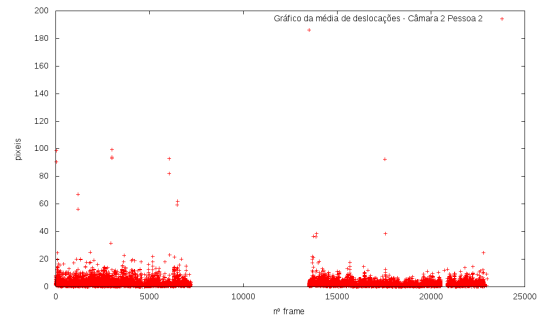


Figura B.2: Gráfico relativo à média de deslocamentos da pessoa com id 2 numa experiência sem indução de conflito (20/12/2017).

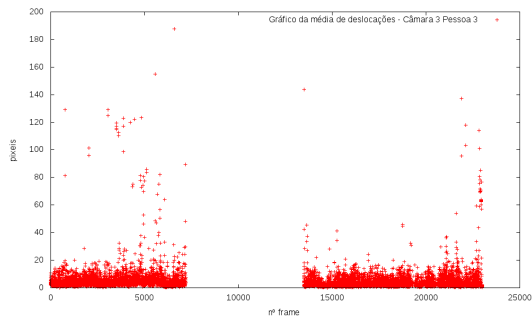


Figura B.3: Gráfico relativo à média de deslocamentos da pessoa com id 3 numa experiência sem indução de conflito (20/12/2017).

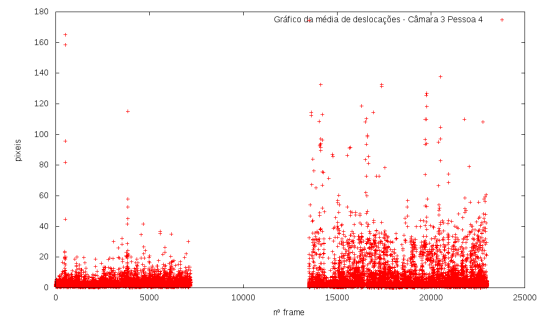


Figura B.4: Gráfico relativo à média de deslocamentos da pessoa com id 4 numa experiência sem indução de conflito (20/12/2017).



## B.2 Média de deslocamentos em experiências com indução de conflito de tarefa

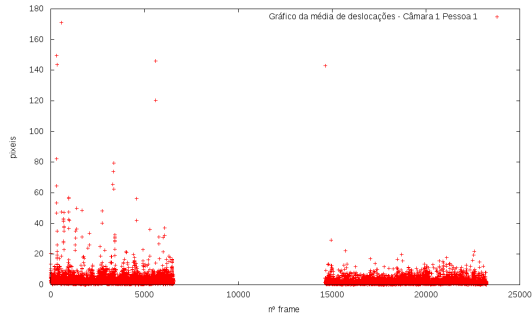


Figura B.5: Gráfico relativo à média de deslocamentos da pessoa com id 1 numa experiência com indução de conflito de tarefa (02/11/2017).

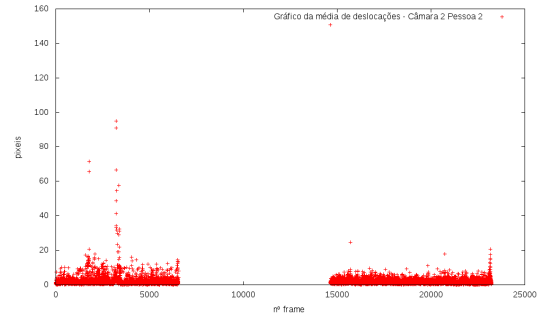


Figura B.6: Gráfico relativo à média de deslocamentos da pessoa com id 2 numa experiência com indução de conflito de tarefa (02/11/2017).

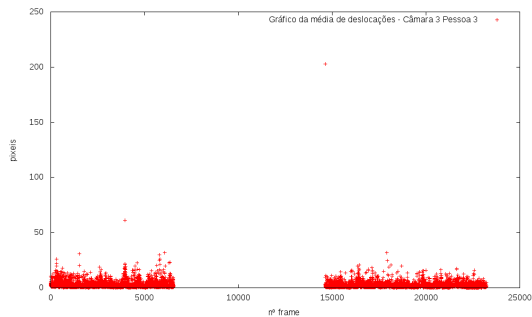


Figura B.7: Gráfico relativo à média de deslocamentos da pessoa com id 3 numa experiência com indução de conflito de tarefa (02/11/2017).

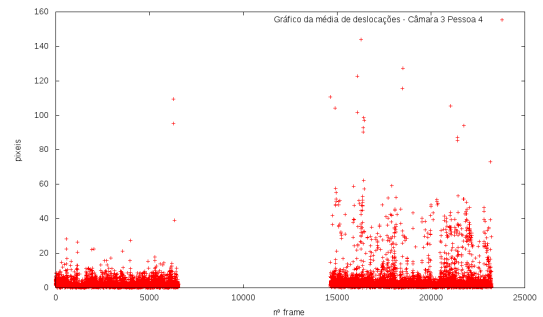


Figura B.8: Gráfico relativo à média de deslocamentos da pessoa com id 4 numa experiência com indução de conflito de tarefa (02/11/2017).

### B.3 Área ocupada em experiências sem indução de conflito

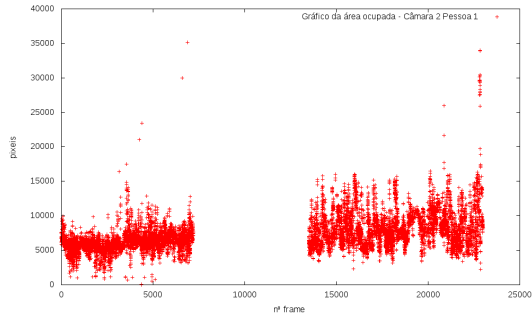


Figura B.9: Gráfico relativo à área ocupada pela pessoa com id 1 numa experiência sem indução de conflito (20/12/2017).

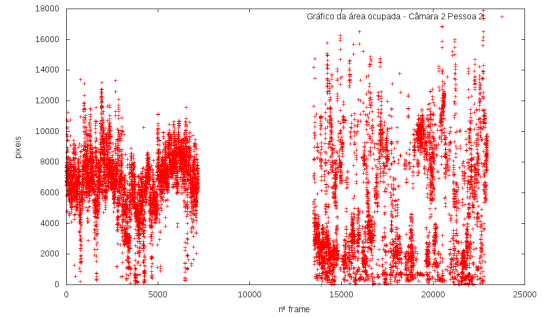


Figura B.10: Gráfico relativo à área ocupada pela pessoa com id 2 numa experiência sem indução de conflito (20/12/2017).

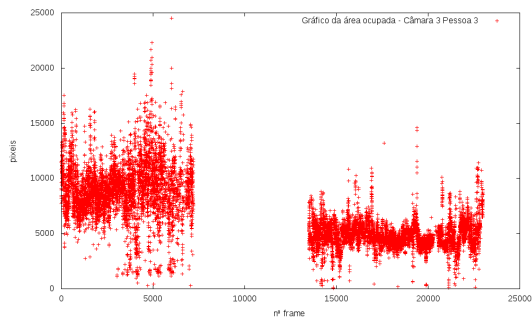


Figura B.11: Gráfico relativo à área ocupada pela pessoa com id 3 numa experiência sem indução de conflito (20/12/2017).

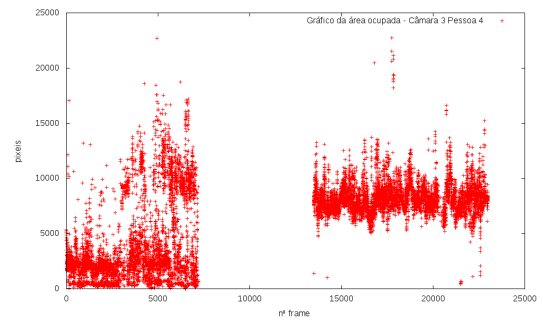


Figura B.12: Gráfico relativo à área ocupada pela pessoa com id 4 numa experiência sem indução de conflito (20/12/2017).

## B.4 Área ocupada em experiências com indução de conflito de tarefa

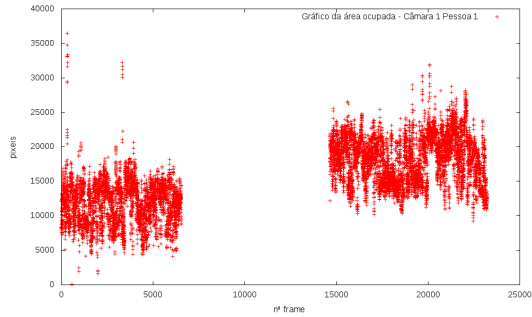


Figura B.13: Gráfico relativo à área ocupada pela pessoa com id 1 numa experiência com indução de conflito de tarefa (02/11/2017).

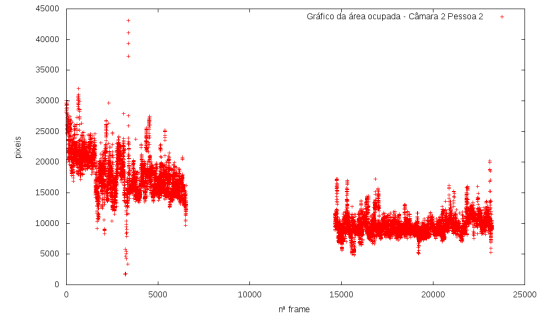


Figura B.14: Gráfico relativo à área ocupada pela pessoa com id 2 numa experiência com indução de conflito de tarefa (02/11/2017).

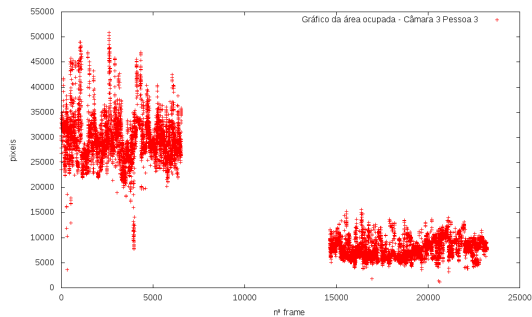


Figura B.15: Gráfico relativo à área ocupada pela pessoa com id 3 numa experiência com indução de conflito de tarefa (02/11/2017).

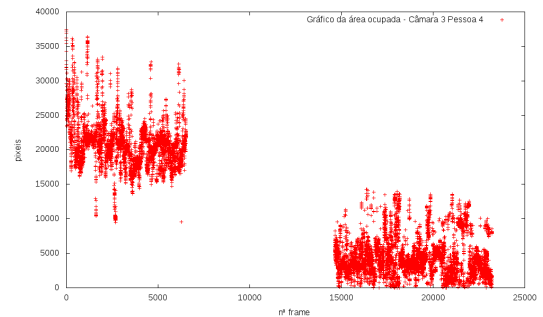


Figura B.16: Gráfico relativo à área ocupada pela pessoa com id 4 numa experiência com indução de conflito de tarefa (02/11/2017).

## B.5 *Heatmaps* relativos a experiências sem indução de conflito

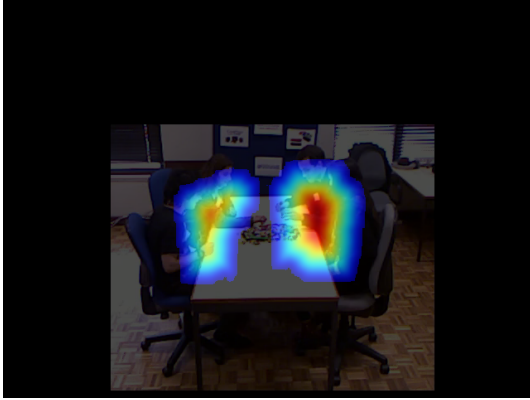


Figura B.17: *Heatmap* relativo à câmara 1 numa experiência sem indução de conflito (20/12/2017).

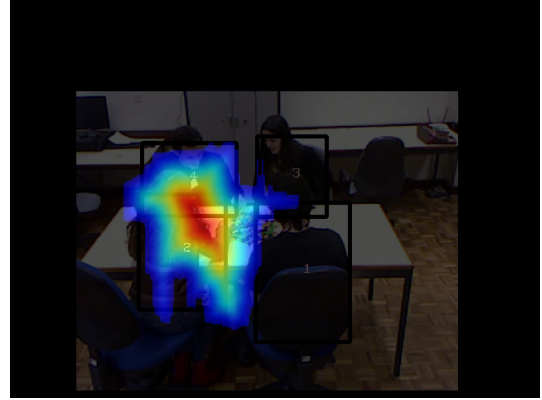


Figura B.18: *Heatmap* relativo à câmara 2 numa experiência sem indução de conflito (20/12/2017).

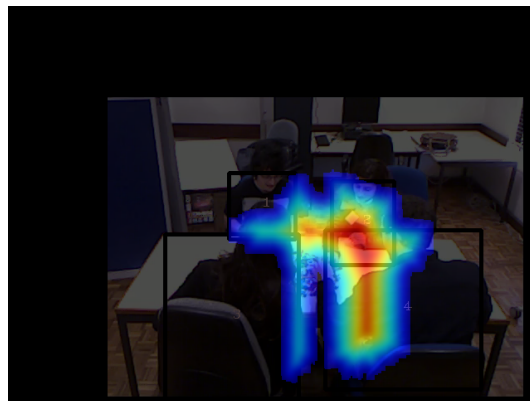


Figura B.19: *Heatmap* relativo à câmara 3 numa experiência sem indução de conflito (20/12/2017).

## B.6 *Heatmaps* relativos a experiências com indução de conflito de tarefa



Figura B.20: *Heatmap* relativo à câmara 1 numa experiência com indução de conflito de tarefa (02/11/2017).

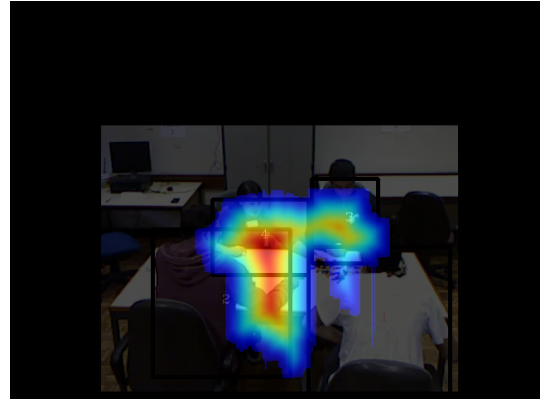


Figura B.21: *Heatmap* relativo à câmara 2 numa experiência com indução de conflito de tarefa (02/11/2017).

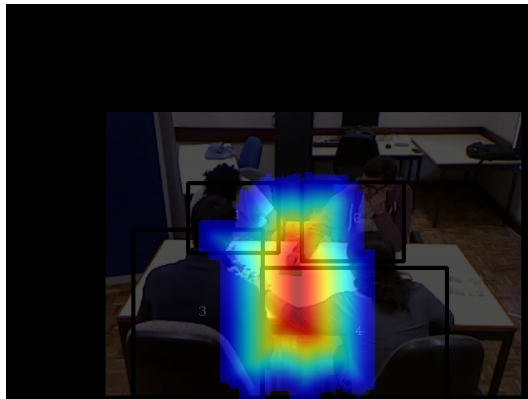


Figura B.22: *Heatmap* relativo à câmara 3 numa experiência com indução de conflito de tarefa (02/11/2017).